

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
25 July 2002 (25.07.2002)

PCT

(10) International Publication Number
WO 02/057450 A2

(51) International Patent Classification⁷: **C12N 15/12**,
C07K 14/47, 16/18, C12N 15/11, G01N 33/68, 33/574,
C12Q 1/68, A61K 31/713, 39/395, 38/17, 31/7088, A61P
9/10, 35/00, 3/10, A01K 67/027

(21) International Application Number: PCT/US01/48922

(22) International Filing Date:
29 November 2001 (29.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/253,834 29 November 2000 (29.11.2000) US
60/250,926 30 November 2000 (30.11.2000) US
60/264,180 25 January 2001 (25.01.2001) US
60/313,656 20 August 2001 (20.08.2001) US
60/327,456 5 October 2001 (05.10.2001) US
Not furnished 28 November 2001 (28.11.2001) US

(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:
US Not furnished (CIP)
Filed on 28 November 2001 (28.11.2001)

(71) Applicant (for all designated States except US): **CURA-
GEN CORPORATION** [US/US]; 555 Long Wharf Drive,
11th floor, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **EDINGER, Shlomit**
[US/US]; 766 Edgewood Avenue, New Haven, CT 06515
(US). **MACDOUGALL, John, R.** [CA/US]; 117 Russell
Street, Hamden, CT 06515 (US). **MILLET, Isabelle**
[FR/US]; 74 Carrington Avenue, Milford, CT 06460 (US).
ELLERMAN, Karen [US/US]; 87 Montoya Drive, Bran-
ford, CT 06405 (US). **STONE, David, J.** [US/US]; 223
Whitethorn Drive, Guilford, CT 06437 (US). **GERLACH,
Valerie** [US/US]; 18 Rock Pasture Road, Branford, CT
06405 (US). **GROSSE, William, M.** [US/US]; 15 Rice
Terrace Road, Apartment C, Branford, CT 06405 (US).
ALSOBROOK, John, P., II [US/US]; 60 Lake Drive,
Madison, CT 06443 (US). **LEPLEY, Denise, M.** [US/US];
51 Church Street, Branford, CT 06405 (US). **RIEGER,
Daniel** [US/US]; 555 Long Wharf Drive, 11th Floor,
New Haven, CT 06405 (US). **BURGESS, Catherine, E.**

[US/US]; 90 Carriage Hill Drive, Weathersfield, CT 06109
(US). **CASMAN, Stacie, J.** [US/US]; 17 Peck Street,
North Haven, CT 06473 (US). **SPYTEK, Kimberly, A.**
[US/US]; 28 Court Street #1, New Haven, CT 06511 (US).
BOLDOG, Ferenc, L. [HU/US]; 1687 Hartford Turn-
pike, North Haven, CT 06473 (US). **LI, Li** [CN/US]; 56
Jerimoth Drive, Branford, CT 06405 (US). **PADIGARU,
Muralindhara** [IN/US]; 71 Hampton Park, Branford, CT
06405 (US). **MISHRA, Vishnu** [IN/US]; 3832 SW 93rd
Terrace, Gainesville, FL 32608 (US). **PATTURAJAN,
Meera** [IN/US]; 45 Harrison Avenue, Apartment 1C,
Branford, CT 06405 (US). **SHENOY, Suresh** [IN/US]; 15
Millwood Drive, Branford, CT 06405 (US). **RASTELLI,
Luca** [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437
(US). **TCHERNEV, Velizar, T.** [BG/US]; 45 Jefferson
Road, #3-12, Branford, CT 06405 (US). **VERNET,
Corine, A., M.** [FR/US]; 1739 Foxon Road, Apartment
L6, North Branford, CT 06471 (US). **ZERHUSEN,
Bryan, D.** [US/US]; 337 Monticello Drive, Branford, CT
06405 (US). **MALYANKAR, Uriel, M.** [IN/US]; 229
Branford Road, # 330, Branford, CT 06405 (US). **GUO,
Xiyojia** [CN/US]; 713 Robert Frost Drive, Branford, CT
06405 (US). **MILLER, Charles, E.** [US/US]; 98 Saddle
Hill Drive, Guilford, CT 06437 (US). **GANGOLLI, Esha,
A.** [IN/US]; 31 Strawberry Hill Road, Madison, CT 06443
(US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris,
Glovsky and Popeo, P.C., One Financial Center, Boston,
MA 02111 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN,
YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).

[Continued on next page]

(54) Title: PROTEINS AND NUCLEIC ACIDS ENCODING SAME

(57) Abstract: Disclosed herein are nucleic acid sequences that encode novel polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies, which immunospecifically-bind to the polypeptide, as well as derivatives, variants, mutants, or fragments of the aforementioned polypeptide, polynucleotide, or antibody. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.

WO 02/057450 A2



Published:

- without international search report and to be republished upon receipt of that report
- entirely in electronic form (except for this front page) and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

PROTEINS AND NUCLEIC ACIDS ENCODING SAME

FIELD OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded thereby.

BACKGROUND OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded therefrom. More specifically, the invention relates to nucleic acids encoding cytoplasmic, nuclear, membrane bound, and secreted polypeptides, as well as vectors, host cells, antibodies, and recombinant methods for producing these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of nucleic acid sequences encoding novel polypeptides. The novel nucleic acids and polypeptides are referred to herein as NOVX, or NOV1, NOV2, NOV3, NOV4, NOV5, NOV6, NOV7, NOV8, NOV9, NOV10, NOV11, and NOV12 nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "NOVX" nucleic acid or polypeptide sequences.

In one aspect, the invention provides an isolated NOVX nucleic acid molecule encoding a NOVX polypeptide that includes a nucleic acid sequence that has identity to the nucleic acids disclosed in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31. In some embodiments, the NOVX nucleic acid molecule will hybridize under stringent conditions to a nucleic acid sequence complementary to a nucleic acid molecule that includes a protein-coding sequence of a NOVX nucleic acid sequence. The invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof. For example, the nucleic acid can encode a polypeptide at least 80% identical to a polypeptide comprising the amino acid sequences of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, and 32. The nucleic acid can be, for example, a genomic DNA fragment or a cDNA molecule that includes the nucleic acid sequence of any of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31.

Also included in the invention is an oligonucleotide, *e.g.*, an oligonucleotide which includes at least 6 contiguous nucleotides of a NOVX nucleic acid (*e.g.*, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31) or a complement of said oligonucleotide.

Also included in the invention are substantially purified NOVX polypeptides (SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, and 32). In certain embodiments, the NOVX polypeptides include an amino acid sequence that is substantially identical to the amino acid sequence of a human NOVX polypeptide.

5 The invention also features antibodies that immunoselectively bind to NOVX polypeptides, or fragments, homologs, analogs or derivatives thereof.

In another aspect, the invention includes pharmaceutical compositions that include therapeutically- or prophylactically-effective amounts of a therapeutic and a pharmaceutically-acceptable carrier. The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide,
10 or an antibody specific for a NOVX polypeptide. In a further aspect, the invention includes, in one or more containers, a therapeutically- or prophylactically-effective amount of this pharmaceutical composition.

In a further aspect, the invention includes a method of producing a polypeptide by culturing a cell that includes a NOVX nucleic acid, under conditions allowing for expression
15 of the NOVX polypeptide encoded by the DNA. If desired, the NOVX polypeptide can then be recovered.

In another aspect, the invention includes a method of detecting the presence of a NOVX polypeptide in a sample. In the method, a sample is contacted with a compound that selectively binds to the polypeptide under conditions allowing for formation of a complex
20 between the polypeptide and the compound. The complex is detected, if present, thereby identifying the NOVX polypeptide within the sample.

The invention also includes methods to identify specific cell or tissue types based on their expression of a NOVX.

Also included in the invention is a method of detecting the presence of a NOVX
25 nucleic acid molecule in a sample by contacting the sample with a NOVX nucleic acid probe or primer, and detecting whether the nucleic acid probe or primer bound to a NOVX nucleic acid molecule in the sample.

In a further aspect, the invention provides a method for modulating the activity of a NOVX polypeptide by contacting a cell sample that includes the NOVX polypeptide with a
30 compound that binds to the NOVX polypeptide in an amount sufficient to modulate the activity of said polypeptide. The compound can be, *e.g.*, a small molecule, such as a nucleic acid, peptide, polypeptide, peptidomimetic, carbohydrate, lipid or other organic (carbon containing) or inorganic molecule, as further described herein.

Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing disorders or syndromes including, *e.g.*, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, hypercoagulation, hemophilia, idiopathic thrombocytopenic purpura, heart failure, secondary pathologies caused by heart failure and hypertension, hypotension, angina pectoris, myocardial infarction, tuberous sclerosis, scleroderma, transplantation, autoimmune disease, lupus erythematosus, viral/bacterial/parasitic infections, multiple sclerosis, autoimmune disease, allergies, immunodeficiencies, graft versus host disease, asthma, emphysema, ARDS, inflammation and modulation of the immune response, viral pathogenesis, aging-related disorders, Th1 inflammatory diseases such as rheumatoid arthritis, multiple sclerosis, inflammatory bowel diseases, AIDS, wound repair, obesity, diabetes, endocrine disorders, anorexia, bulimia, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic, renal tubular acidosis, IgA nephropathy, nephrological diseases, hypercalcaemia, Lesch-Nyhan syndrome, Von Hippel-Lindau (VHL) syndrome, trauma, regeneration (in vitro and in vivo), Hirschsprung's disease, Crohn's Disease, appendicitis, endometriosis, laryngitis, psoriasis, actinic keratosis, acne, hair growth/loss, alopecia, pigmentation disorders, myasthenia gravis, alpha-mannosidosis, beta-mannosidosis, other storage disorders, peroxisomal disorders such as Zellweger syndrome, infantile Refsum disease, rhizomelic chondrodysplasia (chondrodysplasia punctata, rhizomelic), and hyperpyruvic acidemia, osteoporosis, muscle disorders, urinary retention, Albright Hereditary Osteodystrophy, ulcers, Alzheimer's disease, stroke, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, behavioral disorders, addiction, anxiety, pain, neuroprotection, Stroke, Aphakia, neurodegenerative disorders, neurologic disorders, developmental defects, conditions associated with the role of GRK2 in brain and in the regulation of chemokine receptors, encephalomyelitis, anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, Gilles de la Tourette syndrome, leukodystrophies, cancers, breast cancer, CNS cancer, colon cancer, gastric cancer, lung cancer, melanoma, ovarian cancer, pancreatic cancer, kidney cancer, colon cancer, prostate cancer, neuroblastoma, and cervical cancer, Neoplasm; adenocarcinoma, lymphoma; uterus cancer, benign prostatic hypertrophy, fertility, control of growth and development/differentiation related functions such as but not limited

maturation, lactation and puberty, reproductive malfunction, and/or other pathologies and disorders of the like.

The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or a NOVX-specific antibody, or biologically-active derivatives or fragments thereof.

5 For example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding NOVX may be useful in
10 gene therapy, and NOVX may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

 The invention further includes a method for screening for a modulator of disorders or
15 syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The method includes contacting a test compound with a NOVX polypeptide and determining if the test compound binds to said NOVX polypeptide. Binding of the test compound to the NOVX polypeptide indicates the test compound is a modulator of activity, or of latency or predisposition to the aforementioned disorders or syndromes.

20 Also within the scope of the invention is a method for screening for a modulator of activity, or of latency or predisposition to disorders or syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like by administering a test compound to a test animal at increased risk for the aforementioned disorders or syndromes. The test animal expresses a recombinant polypeptide encoded by a
25 NOVX nucleic acid. Expression or activity of NOVX polypeptide is then measured in the test animal, as is expression or activity of the protein in a control animal which recombinantly-expresses NOVX polypeptide and is not at increased risk for the disorder or syndrome. Next, the expression of NOVX polypeptide in both the test animal and the control animal is compared. A change in the activity of NOVX polypeptide in the test animal relative to the
30 control animal indicates the test compound is a modulator of latency of the disorder or syndrome.

 In yet another aspect, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a NOVX polypeptide, a NOVX nucleic acid, or both, in a subject (*e.g.*, a human subject). The method includes measuring the

amount of the NOVX polypeptide in a test sample from the subject and comparing the amount of the polypeptide in the test sample to the amount of the NOVX polypeptide present in a control sample. An alteration in the level of the NOVX polypeptide in the test sample as compared to the control sample indicates the presence of or predisposition to a disease in the subject. Preferably, the predisposition includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. Also, the expression levels of the new polypeptides of the invention can be used in a method to screen for various cancers as well as to determine the stage of cancers.

In a further aspect, the invention includes a method of treating or preventing a pathological condition associated with a disorder in a mammal by administering to the subject a NOVX polypeptide, a NOVX nucleic acid, or a NOVX-specific antibody to a subject (*e.g.*, a human subject), in an amount sufficient to alleviate or prevent the pathological condition. In preferred embodiments, the disorder, includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

In yet another aspect, the invention can be used in a method to identify the cellular receptors and downstream effectors of the invention by any one of a number of techniques commonly employed in the art. These include but are not limited to the two-hybrid system, affinity purification, co-precipitation with antibodies or other specific-interacting molecules.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences and their encoded polypeptides. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX

polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table A provides a summary of the NOVX nucleic acids and their encoded polypeptides.

5

TABLE A. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	Homology
1a	Gmba58o1_A_da1	1	2	Transmembrane receptor UNC5H2-like
1b	Gmba58o1_A	3	4	Transmembrane receptor UNC5H2-like
2a	SC126422078_A	5	6	Tyrosine Phosphatase Precursor-like
2b	CG50718-02	7	8	Glomerular Mesangial Cell Receptor Protein Tyrosine Phosphatase Precursor like
2c	CG50718-05	9	10	Glomerular Mesangial Cell Receptor Protein Tyrosine Phosphatase Precursor like
3	134899552_EXT	11	12	Human homolog of the <i>Drosophila</i> pecanex-like
4	SC140515441_A	13	14	Aurora-related kinase 1-like
5	SC44326718_A	15	16	26S protease regulatory subunit 4-like
6	GMAC073364_Ada1	17	18	Mitsugumin29-like
7	106973211_EXT	19	20	Wnt-15-like
8	88091010-EXT	21	22	Wnt-14-like
9	AC069250_28_da1	23	24	Beta-adrenergic receptor kinase-like
10	AC058790_da25	25	26	Alpha-mannosidase-like
11a	GM57107065_da1	27	28	Clq-related factor-like
11b	CG54503-02	29	30	Clq-related factor-like
12	SC132340676_A	31	32	Plexin 1-like

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

NOV1 is homologous to the transmembrane receptor UNC5H2-like family of proteins. Thus, NOV1 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example; cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus,

pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, Lesch-Nyhan syndrome, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, cancers, and/or other pathologies and disorders. Also since this gene is expressed at a measurably higher level in several cancer cell lines (including breast cancer, CNS cancer, colon cancer, gastric cancer, lung cancer, melanoma, ovarian cancer and pancreatic cancer), it may be useful in diagnosis and treatment of these cancers.

NOV2 is homologous to the protein tyrosine phosphatase precursor-like family of proteins. Thus NOV2 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example; cancer, kidney cancer, trauma, regeneration (in vitro and in vivo), viral/bacterial/parasitic infections, nephrological diseases including diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, Lesch-Nyhan syndrome, Hirschsprung's disease, Crohn's Disease, appendicitis, and/or other pathologies and disorders.

NOV3 is homologous to the Human homolog of the *Drosophila* pecanex family of proteins. Thus NOV3 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example; cancer, trauma, regeneration (in vitro and in vivo), viral/bacterial/parasitic infections, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, multiple sclerosis, scleroderma, obesity, endometriosis, fertility, hypercoagulation, autoimmune disease, allergies, immunodeficiencies, transplantation, hemophilia, idiopathic thrombocytopenic purpura, graft versus host disease, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, systemic lupus erythematosus, asthma, emphysema, ARDS, laryngitis, psoriasis, actinic keratosis, acne, hair growth/loss,

alopecia, pigmentation disorders, endocrine disorders, diabetes, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, Lesch-Nyhan syndrome, and a variety of kidney diseases and/or other pathologies and disorders.

5 NOV4 is homologous to a family of Aurora-related kinase 1-like proteins. Thus, the NOV4 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example: breast, ovarian, colon, prostate, neuroblastoma, and cervical cancer, Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect
10 (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation, Diabetes, Von Hippel-Lindau (VHL) syndrome, Pancreatitis, Alzheimer's disease, Stroke, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia,
15 Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, and Neuroprotection, and/or other pathologies.

 NOV5 is homologous to the 26S protease regulatory subunit 4-like family of proteins. Thus, NOV5 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example:
20 cataract and Aphakia, Alzheimer's disease, neurodegenerative disorders, inflammation and modulation of the immune response, viral pathogenesis, aging-related disorders, neurologic disorders, cancer, and/or other pathologies.

 NOV6 is homologous to the MITSUGUMIN29-like family of proteins. Thus, NOV6 nucleic acids, polypeptides, antibodies and related compounds according to the invention will
25 be useful in therapeutic and diagnostic applications implicated in, for example: muscular dystrophy, Lesch-Nyhan syndrome, myasthenia gravis, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial
30 septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, and other diseases, disorders and conditions of the like. Also since the invention is highly

expressed in one of the lung cancer cell lines (Lung cancer NCI-H522), it may be useful in diagnosis and treatment of this cancer.

NOV7 is homologous to the Wnt-15-like family of proteins. Thus NOV7 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in
5 Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, cancer, developmental defects, and/or other pathologies/disorders.

10 NOV8 is homologous to members of the Wnt-14-like family of proteins. Thus, the NOV8 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example; Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-
15 Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, cancer, developmental defects, and/or other pathologies/disorders.

NOV9 is homologous to the beta adrenergic receptor kinase-like family of proteins. Thus, NOV9 nucleic acids and polypeptides, antibodies and related compounds according to
20 the invention will be useful in therapeutic and diagnostic applications implicated in, for example: heart failure, hypertension, secondary pathologies caused by heart failure and hypertension, and other diseases, disorders and conditions of the like. Additionally, the compositions of the present invention may have efficacy for treatment of patients suffering from conditions associated with the role of GRK2 in brain and in the regulation of chemokine
25 receptors.

NOV10 is homologous to the alpha-mannosidase-like family of proteins. Thus, NOV10 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example: alpha-mannosidosis, beta-mannosidosis, other storage disorders, peroxisomal disorders such as
30 zellweger syndrome, infantile refsum disease, rhizomelic chondrodysplasia (chondrodysplasia punctata, rhizomelic), and hyperpipecolic acidemia and other diseases, disorders and conditions of the like, and/or other pathologies/disorders.

NOV11 is homologous to the C1q-related factor-like family of proteins. Thus, NOV11 nucleic acids and polypeptides, antibodies and related compounds according to the invention

will be useful in therapeutic and diagnostic applications implicated in, for example: Th1 inflammatory diseases such as rheumatoid arthritis, multiple sclerosis, inflammatory bowel diseases and psoriasis, lupus erythematosus and glomerulonephritis, control of growth and development/differentiation related functions such as but not limited maturation, lactation and puberty, osteoporosis, obesity, aging and reproductive malfunction and hence could be used in treatment and/or diagnosis of these disorders.

NOV12 is homologous to the Plexin-1 like family of proteins. Thus, NOV12 nucleic acids and polypeptides, antibodies, and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example: AIDS, cancer therapy, treatment of Neurologic diseases, Brain and/or autoimmune disorders like encephalomyelitis, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, endocrine diseases, muscle disorders, inflammation and wound repair, bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome, and/or other pathologies/disorders.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit, *e.g.*, neurogenesis, cell differentiation, cell proliferation, hematopoiesis, wound healing and angiogenesis.

Additional utilities for the NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

NOV1

NOV1 includes three novel transmembrane receptor UNC5H2-like proteins disclosed below. The disclosed sequences have been named NOV1a and NOV1b.

NOV1a

A disclosed NOV1a nucleic acid of 2860 nucleotides (also referred to as GMba58o1_A_dal) encoding a transmembrane receptor UNC5H2-like protein is shown in Table 1A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 59-61 and ending with a TGA codon at nucleotides 2858-2860. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 1A. The start and stop codons are in bold letters.

Table 1A. NOV1a nucleotide sequence (SEQ ID NO:1).

```

AGACTGGGGCCAGGGAGACAGCCCTGGGGGAGAGGGCGCCGAACCAGGCGCGGGAGCATGGGGGCCCCGGAG
CGGAGCTCGGGGCGCGCTGCTGCTGGCACTGCTGCTCTGCTGGGACCGAGGCTGAGCCAAGCAGGCACTGA
TTCTGGCAGCGAGGTGCTCCCTGACTCCTTCCCGTCAGCGCCAGCAGAGCCGCTGCCCTACTTCTGCAAGGA
GCCACAGGACGCCTACATTGTGAAGAACAAGCCTGTGGAGCTCCGCTGCCGCGCCTTCCCGCCACACAGAT
CTACTTCAAGTCAACGGCGAGTGGGTGAGCCAGAACCACGTCACACAGGAAGGCCTGGATGAGGCCAC
CGGTCTGCGGGTGCGCGAGGTGCAGATCGAGGTGTGCGGCGAGCAGGTGGAGGAGCTCTTTGGGTGGAGGA
TTACTGGTGCCAGTGGCTGGCCTGGAGCTCCGCGGGCACCAACAGAGTCCGCGAGCCTACGTCGCGCATCGC
CTACCTGCGCAAGAACTTCGATCAGGAGCCTCTGGGCAAGGAGTGGCCCTGGACCATGAGGTTCTCCTGCA
GTGCCGCGCGCGGAGGGGGTGCTGTGGCCGAGGTGGAATGGCTCAAGAATGAGGATGTCATCGACCCAC
CCAGGACACCAACTTCCCTGCTCACCATCGACCACAACCTCATCATCCGCGAGGCCCGCTGTGCGACACTGC
CAACTATACCTGCGTGGCCAGAAGCATCGTGGCCAAACGCCGAGCAACCACTGCCACCGTCATCGTCTACGT
GAATGGCGGCTGGTCCAGCTGGGCAGAGTGGTCACCCTGCTCCAACCGCTGTGGCCGAGGCTGGCAGAAGCG
CACCCGACCTGCACCAACCCCGCTCCACTCAACGGAGGGGCTTCTGCGAGGGCCAGGCATTCCAGAAGAC
CGCCTGCACCACCATCTGCCAGTCGATGGGGCGTGGACGGAGTGGAGCAAGTGGTCAGCCTGCAGCACTGA
GTGTGCCCACTGGCGTAGCCGCGAGTGATGGCGCCGCCACCCAGAACGAGGCGCTGACTGCAGCGGAC
GCTGCTCGACTCTAAGAACTGCACAGATGGGCTGTGCATGCAACTGGAGGCTCAGGGGATGCGGCGCTGTA
TGCGGGGCTCGTGGTGGCCATCTTCGTGGTGTGGCAATCCTCATGGCGGTGGGGGTGGTGGTGTACCGCCG
CAACTGCCGTGACTTCGACACAGACATCACTGACTCATCTGCTGCCCTGACTGGTGGTTCACCCCGTCAA
CTTTAAGACGGCAAGGCCAGTAACCCGAGCTCCTACACCCCTCTGTGCCTCTGACCTGACAGCCAGCGC
CGGCATCTACCGCGGACCCGTGTATGCCCTGCAGGACTCCACCGACAAATCCCATGACCAACTCTCCTCT
GCTGGACCCCTTACCCAGCCTTAAGGTCAAGGTCTACAGCTCCAGCACCACGGGCTCTGGGCCAGGCCTGGC
AGATGGGGCTGACCTGCTGGGGGTCTTGCCGCTGGGCACATACCTAGCGATTTGCGCCGGGACACCCACTT
CCTGCACCTGCGCAGCGCCAGCCTCGGTTCCAGCAGCTCTTGGGCTGCCCGAGACCCAGGGAGCAGCT
CAGCGGCACCTTTGGCTGCTGGGTGGGAGGCTCAGCATCCCGGCACAGGGGTGAGCTTGTGGTGGCCAA
TGGAGCCATTCCCAGGGCAAGTCTACGAGATGTATCTACTCATCAACAGGCAGAAAGTACCCTGCCGCT
TTCAGAAAGGACCCAGACAGTATTGAGCCCTCGGTGACCTGTGGACCCACAGGCCTCCTGCTGTGCCGCC
CGTCATCCTCACCATGCCCCACTGTGCCGAAGTCAGTGCCCGTGAATCTTTCAGCTCAAGACCCAGGC
CCACAGGGCCTTGGCTGCTGGGTGGGAGGCTGAGTACCCCTGAGGAGACCTGAACACACCCCTGCTACTGCCAGCT
GGAGCCCAGGGCCTGTACATCCTGCTGGACAGCTGGGCACCTACGTGTTACGGGCGAGTCTATTCCCG
CTCAGCAGTCAAGCGGCTCCAGCTGGCCGTCTTCGCCCCCGCCCTCTGCACCTCCCTGGAGTACAGCCTCCG
GGTCTACTGCCTGGAGGACACGCTGTAGCACTGAAGGAGGTGCTGGAGCTGGAGCGGACTCTGGGCGGATA
CTTGTGTGAGGAGCCGAAACCGCTAATGTTCAAGGACAGTTACCACAACCTGCGCCTCTCCCTCCATGACCT
CCCCATGCCCATTTGGAGGAGCAAGCTGCTGGCCAAATACCAGGAGATCCCTTCTATCACATTGGAGTGG
CAGCCAGAAGGCCCTCCACTGCACTTTCACCCTGGAGAGGCACAGCTTGGCCTCCACAGAGCTCACCTGCAA
GATCTGCGTGCGCAAGTGAAGGGGAGGGCCAGATATTCCAGCTGCATACCACTCTGGCAGAGACACCTGC
TGGCTCCCTGGACACTCTCTGCTCTGCCCTGGCAGCACTGTACCACCCAGCTGGGACCTTATGCCTTCAA
GATCCCACTGTCCATCCGCCAGAAGATATGCAACAGCCTAGATGCCCCCAACTCACGGGGCAATGACTGGCG
GATGTTAGCACAGAAGCTCTCTATGGACCGGTACCTGAATTACTTTGCCACCAAGCGAGCCCCACGGGTGT
GATCCTGGACCTCTGGGAAGCTCTGCAGCAGGACGATGGGGACCTCAACAGCCTGGCGAGTGCCTTGGAGGA
GATGGGCAAGAGTGAGATGCTGGTGGCTGTGGCCACCGGGGACTGCTGA

```

In a search of public sequence databases, the NOV1a nucleic acid sequence, located on chromosome 10 has 1604 of 1895 bases (84%) identical to a transmembrane receptor UNC5H2 mRNA from *Rattus Norvegicus*, (GENBANK-ID: RNU87306). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

In all BLAST alignments herein, the "E-value" or "Expect" value is a numeric indication of the probability that the aligned sequences could have achieved their similarity to the BLAST query sequence by chance alone, within the database that was searched. For example, the probability that the subject ("Sbjct") retrieved from the NOV1a BLAST analysis, e.g., transmembrane receptor UNC5H2 mRNA from *Rattus Norvegicus*, matched the Query NOV1a sequence purely by chance is 0.0. The Expect value (E) is a parameter that describes the number of hits one can "expect" to see just by chance when searching a database of a particular size. It decreases exponentially with the Score (S) that is assigned to a match between two sequences. Essentially, the E value describes the random background noise that exists for matches between sequences.

The Expect value is used as a convenient way to create a significance threshold for reporting results. The default value used for blasting is typically set to 0.0001. In BLAST 2.0, the Expect value is also used instead of the P value (probability) to report the significance of matches. For example, an E value of one assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see one match with a similar score simply by chance. An E value of zero means that one would not expect to see any matches with a similar score simply by chance. See, e.g., <http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/>. Occasionally, a string of X's or N's will result from a BLAST search. This is a result of automatic filtering of the query for low-complexity sequence that is performed to prevent artifactual hits. The filter substitutes any low-complexity sequence that it finds with the letter "N" in nucleotide sequence (e.g., "NNNNNNNNNNNNNN") or the letter "X" in protein sequences (e.g., "XXXXXXXXXX"). Low-complexity regions can result in high scores that reflect compositional bias rather than significant position-by-position alignment. (Wootton and Federhen, *Methods Enzymol* 266:554-571, 1996).

The disclosed NOV1a polypeptide (SEQ ID NO:2) encoded by SEQ ID NO:1 has 933 amino acid residues and is presented in Table 1B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV1a has a signal peptide at the first 26 amino acids and is likely to be localized at the plasma membrane with a certainty of 0.5140. In other embodiments, NOV1a is likely to be localized to the microbody (peroxisome) with a certainty of 0.1064, to the endoplasmic reticulum (membrane) with a certainty of 0.1000, or to the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV1a is between positions 26 and 27: SQA-GT

Table 1B. Encoded NOV1a protein sequence (SEQ ID NO:2).

```

MGARSGARGALLLALLLCWDFRLSQAGTDSGSEVLPSDFPSAPAEPLPYFLQEPQDAYIVKNKPVELRCRAF
PATQIYFKCNGEWSQNDHVTQEGLEDEATGLRVREVQIEVSRQQVEELFGLLEDYWCQCVAWSSAGTTKSRRA
YVRIAYLRKNFDQEPFGKEVPLDHEVLLQCRPPEGVPFAEVEWLKNEVDVIDPTQDTNFLTIDHNLIIRQAR
LSDTANYTCVAKNIVAKRRSTTATVIVYVNGGWSSWAESWPCSNRCGRGWQKRTRTCTNPAPLNGGAFCEGQ
AFQKTACTTICPVDGAWTEWSKWSACSTECAHWSRECMAPPQNGGRDCSGTLLDSKNCTDGLCMQLEASG
DAALYAGLVVAIFVVVAILMAVGVVYRRNCRDFDITDSSAALTGGFHPVNEKTARPSNPQLLHPSVPPD
LTASAGIYRGPVYALQDSTDKIPMTNSPLLDPLPSLKVKVYSSSTTSGSGPLADGADLLGVLPFGTYPSDFA
RDTHFLHLRSASLGSQQLGLPRDPGSSVSGTFGLGGRLSIPGTGVSLVLPNGAIPQGKFYEMYLLINKAE
STLPLSEGTQTVLSPSVTCGPTGLLLCRPVILTMPHCAEVSARDWIFQLKTOAHQGHWEVVTLDDEETLNTF
CYCQLEPRACHILLDQLGTIVFTGESYSRSVAVKRLQLAVFAPALCTSLVSLRVYCLEDTPVALKEVLELER
TLGGYLVVEPKPLMEKDSYHNLRSLHDLPHAHWSKLLAKYQEIIPFYHIWSGSQKALHCTFTLERHSLAST
ELTCKICVRQVEGEQIFQLHTTLAETPAGSLDTLCSAPGSTVTTQLGPYAFKIPLSIRQKICNSLDAENSR
GNDWRMLAQKLSDRYLNYFATKASPTGVILDLWEALQDDGDLNSLASALEEMKSEMLVAVATDGDGDC

```

A search of sequence databases reveals that the NOV1a amino acid sequence has 862 of 945 amino acid residues (91%) identical to, and 897 of 945 amino acid residues (94%) similar to, the 945 amino acid residue 6330415E02RIK protein from *Mus musculus* (Q9D398) (E = 0.0). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV1a is at least expressed in endothelial cells, heart, kidney, adipose, brain (hippocampus), brain (thalamus), cerebral cortex, and the following cancer cell lines: breast cancer, CNS cancer, colon cancer, gastric cancer, lung cancer, melanoma, ovarian cancer and pancreatic cancer at a measurably higher level than the following tissues: adrenal gland, bladder, bone marrow, brain (amygdala), brain (cerebellum), brain (whole), breast, colorectal, liver, lung, lymph node, mammary gland, ovary, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thymus, thyroid gland, trachea, and uterus.

NOV1b

A disclosed NOV1b nucleic acid of 2860 nucleotides (also referred to as CG50126-02) encoding a novel beta-adrenergic receptor kinase-like protein is shown in Table 1C. An open reading frame was identified beginning with an ATG codon at nucleotides 59-61, and ending with a TGA codon at nucleotides 2858-2860. Putative untranslated regions, if any, are located upstream from the initiation codon and downstream from the termination codon.

Table 1C. NOV1b nucleotide sequence (SEQ ID NO:3).

```

AGACTGGGGCCAGGAGACAGCCCTGGGGGAGAGGCGCCCGAACAGGCCGCGGGAACATGGGGGCCCGGAGCGGAGCTC
GGGCGCGCTGCTGCTGGCACTGCTGCTCTGCTGGGACCCGAGGCTGAGCCAAGCAGGCACTGATTCTGGCAGCGAGGTG
CTCCCTGACTCCTTCCCGTCAGCGCCAGCAGAGCCGCTGCCCTACTTCTGCAGGAGCCACAGGACGCTACATTGTGAA
GAACAAGCCTGTGGAGCTTCGCTGCCGCGCCTTCCCCGCCACACAGATCTACTTCAAGTGCAACGCGAGTGGGTGAGCC
AGAACGACCACGTACACAGGAAGGCTGGATGAGGCCACCGGCTGCGGGTGGCGAGGTGCAGATCGAGGTGTGCGGG
CAGCAGGTGGAGGAGCTCTTTGGGCTGGAGGATTACTGGTGCCAGTGGCTGGCCTGGAGCTCCGCAGGCAACCAAGAG
TCGCCGAGCCTACGTCGGCATCGCCTACCTGCGCAAGAACTTCGATCAGGAGCCTCTGGGCAAGGAGGTGCCCCCTGGACC
ATGAGGTTCTCCTGCAGTGCCGCCCCGCGAGGGGGTGCCTGTGGCCGAGGTGGAATGGCTCAAGAATGAGGATGTCTATC
GACCCACCCAGGACACCAACTTCTGCTCACCATCGACCACAACCTCATCATCCGCCAGGCCCGCCTGTGCGGACACTGC
CAACTATACCTGCGTGGCCAAAGAACATCGTGCCAAACGCCGGAGCACCCTGCCACCGTCTCGTCTACGTGAATGGCG
GCTGGTCCAGCTGGGCGAGTGGTCAACCTGCTCCAACCGCTGTGGCCGAGGCTGGCAGAAGCGCACCCGGACCTGCACC
AACCCCGTCCACTCAACGAGGGGCTTCTGCGAGGGCCAGGCAATCCAGAAGACCGCTGCACCACTATGCCCAGT
CGATGGGGGCTGGAGCGAGTGGAGCAAGTGGTGCAGCTGCAGCACTGAGTGTGCCCACTGGCGTAGCCGCGAGTGCATGG
CGCCCCACCCAGAACGAGGCGGTGACTGCAGCGGACGCTGCTCGACTTAAGAAGTGCACAGATGGGCTGTGCATG
CAACTGGAGGCTCAGGGGATGCGGCGCTGTATGCGGGGCTCGTGGTGGCCATCTTCGTGGTCTGGCAATCCTCATGGC
GGTGGGGTGGTGGTGTACCGCGCAACTGCCGTGACTTCGACACAGACATCACTGACTCATCTGCTGCCCTGACTGGTG
GTTTCCACCCCGTCAACTTTAAGACGCGCAAGGCCAGTAACCCGAGCTCCTACACCCCTCTGTGCCTCTGACCTGACA
GCCAGCGCCGCGCATCTACCGCGGACCCGTGTATGCCCTGCAGGACTCCACCGACAAAATCCCATGACCAACTCTCCTCT
GCTGGACCCCTTACCAGCCTTAAGGTCAAGGTCTACAGCTCCAGCACCACGGGCTCTGGGCCAGGCTGGCAGATGGGG
CTGACCTGCTGGGGTCTTGCCGCTGGCACATACCTAGCGATTTCGCCCGGGACACCCACTTCTGCACCTGCGCAGC
GCCAGCCTCGGTTCACAGCAGCTCTTGGGCTGCCCCGAGACCCAGGAGCAGCGTCAAGCGGACCTTTGGCTGCCTGGG
TGGGAGGCTCAGCATCCCCGGCACAGGGGTGAGCTTGTGGTGGCCCAATGGAGCCATTCCCAGGGCAAGTCTACGAGA
TGATCTACTCATCAACAAGGCAGAAAGTACCTGCGCTTTTCAAGAGGACCCAGACAGTATTGAGCCCTCGGTGACC
TGTGGACCCACAGGCTCCTGCTGTGCCGCCCCGTCTCTCACCATGCCCCACTGTGCCGAAGTCACTGCCCGTCACTG
GATCTTTCACTCAAGACCCAGGCCACAGGGCCACTGGGAGGAGTGGTGACCTGGATGAGGAGACCTGAACACAC
CCTGCTACTGCCAGTGGAGCCAGGGCTGTACATCCTGCTGGACAGCTGGGCACCTACGTGTTACGGGCGAGTCC
TATTCGCGCTCAGCAGTCAAGCGGCTCCAGCTGCGCGTCTTCCGCCCGCCCTCTGCACCTCCCTGGAGTACAGCTCCG
GGTCTACTGCTGGAGGACAGCCTGTAGCACTGAAGGAGGTGCTGGAGCTGGAGCGGACTCTGGGCGGATACTTGGTGG
AGGAGCCGAAACCGTAATGTTCAAGGACAGTTACCAACAGCTGCGCCTCTCCCTCCATGACCTCCCCATGCCCCATGG
AGGAGCAAGCTGCTGGCCAAATACAGGAGATCCCCCTTCTATACATTTGGAGTGGCAGCCAGAAGGCCCTCCACTGCAC
TTTCAACCCTGGAGAGGCACAGCTTGGCCTCCACAGAGCTCACCTGCAAGATCTGCGTGGGCAAGTGGAGGGGAGGGCC
AGATATTCCAGCTGCATACCACTCTGGCAGAGACACCTGCTGGCTCCCTGGACACTCTCTGCTCTGCCCTGGCAGCACT
GTCACCAACAGCTGGGACCTTATGCCCTCAAGATCCCACTGTCCATCCGCCAGAAGATATGCAACAGCCTAGATGCCCC
CAACTCAGCGGGCAATGACTGGCGGATGTTAGCACAGAAGCTCTATGGACCGGTACCTGAATTACTTTGCCACCAAG
CGAGCCCCACGGGTGTGATCCTGGACCTCTGGGAAGCTCTGCAGCAGGACGATGGGGACCTCAACAGCCTGGCGAGTGGC
TTGAGGAGATGGGCAAGAGTGAGATGCTGTTGGCTGTGGCCACCGACGGGACTGCTGA

```

In a search of public sequence databases, the NOV1b nucleic acid sequence, located on chromosome 10 has 1604 of 1895 bases (84%) identical to a gb:GENBANK-

ID:RNU87306|acc:U87306.1 mRNA from *Rattus norvegicus* (*Rattus norvegicus*

5 transmembrane receptor Unc5H2 mRNA, complete cds). (E = 0.0) Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV1b polypeptide (SEQ ID NO:4) encoded by SEQ ID NO:3 has 933 amino acid residues and is presented in Table 1D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV1b has a signal peptide at the first 26
 10 amino acids and is likely to be localized at the plasma membrane with a certainty of 0.5140. In other embodiments, NOV1b is likely to be localized to the microbody (peroxisome) with a certainty of 0.1064, to the endoplasmic reticulum (membrane) with a certainty of 0.1000, or to the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV1b is between positions 26 and 27: SQA-GT

Table 1D. Encoded NOV1b protein sequence (SEQ ID NO:4).

```

MGARSGARGALLLALLLCWDFRLSQAGTDSGSEVLDPDSFPSAPAEPLPYFLQEPQDAYIVKNKPVELRCRAFPATQIYFK
CNGEWWVSQNDHVTQEGLDDEATGLRVREVQIEVSRQQVEELFGLLEDYWCQCVAWSSAGTTKSRRAYVRIAYLRKNFDQEP
L
GKEVPLDHEVLLQCRPPEGVPFAEVEWLKNEVDIDPTQDTNFLTIDHNLIIRQARLSDTANYTCVAKNIVAKRRSTTAT
VIVYVNGGWSWAEWSPCSNRCGRGWQKRTRTCTNPAPLNGGAFCEGQAFQKTACTTICPVDGAWTEWSKWSACSTECAL
WRSRECMAPPQNGGRDCSGTLLDSKNCTDGLCMQLEASGDAALYAGLVVAIFVVVAAILMAVGVVVYRRNCRDFDITD
SSAALTGGFHPVNFKTARPSNPQLLHPSVPPDLTASAGIYRGPVYALQDSTDKIPMTNSPLDPLPSLVKVYSSSTTGS
GPGGLADGADLLGVLPFGTYPSPDFARDTHFLHLSASLGSQQLLGLPRDPGSSVSGTFGCLGGRSLIPGTGVSLVFPNGAI
PQKGFYEMYLLINKAESTLPLSEGTQTVLSPSVTCGPTGLLLCRPVILTMPHCAEVSARDWIFQLKTAHQGHWEVVTL
DEETLNTPCYCQLEPRACHILLDQLGTYVFTGESYSRSVAVKRLQLAVFAPALCTSLEYSLRVYCLEDTPVALKEVELEL
TLGGYLVVEPKPLMFKDSYHNLRSLHDLPHAHWRSKLLAKYQEIIPFYHIWGSQKALHCTFTLERHSLASTELTCKICV
RQVEGEGQIFQLHTTIAETPAGSLDTLCSAPGSTVTTQLGPYAFKIPLSIRQKICNSLDAPNSRGNDWRMLAQKLSMDRY
LNYFATKASPTGVILDLEALQDDGDLNSLASALEEMGKSEMLVAVATDGDG

```

A search of sequence databases reveals that the NOV1b amino acid sequence has 862 of 945 amino acid residues (91%) identical to, and 893 of 945 amino acid residues (94%) similar to, the 945 amino acid residue ptnr:SPTREMBL-ACC:O08722 protein from *Rattus norvegicus* (Rat) (Transmembrane Receptor UNC5H2) (E = 0.0). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV1b is expressed in at least adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, and/or RACE sources.

The disclosed NOV1a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 1E.

Table 1E. BLAST results for NOV1a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
ptnr:SPTREMBL- ACC:Q9D398	6330415E02RIK PROTEIN - Mus musculus (Mouse)	945	862/945 (91%)	897/945 (94%)	0.0
ptnr:SPTREMBL- ACC:O08722	TRANSMEMBRANE RECEPTOR UNC5H2	945	862/945 (91%)	893/945 (94%)	0.0
ptnr:SPTREMBL- ACC:O08747	UNC-5 HOMOLOG (C. ELEGANS)	931	610/929 (65%)	723/929 (77%)	0.0
ptnr:SPTREMBL- ACC:O95185	TRANSMEMBRANE RECEPTOR UNC5C - Homo sapiens	931	598/929 (64%)	718/929 (77%)	0.0

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 1F. In the ClustalW alignment of the NOV1 proteins, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 1F. ClustalW Analysis of NOV1

10	1) NOV1a (SEQ ID NO:2)
	2) NOV1b (SEQ ID NO:4)
	3) ptmr: 6330415E02RIK PROTEIN - Mus musculus (Mouse) (SEQ ID NO:33)
	4) ptmr: TRANSMEMBRANE RECEPTOR UNC5H2 (SEQ ID NO:34)
	5) ptmr: UNC-5 HOMOLOG (C. ELEGANS) (SEQ ID NO:35)
15	6) ptmr: TRANSMEMBRANE RECEPTOR UNC5C - Homo sapiens (SEQ ID NO:36)
20	NOV1a -----RCARSGA--KGALLLALLLCWDIFRISAGTDCGSE-----VLPDSFSPAPAEPLFYFLQKIQDAYIVNKKPVE 66 NOV1b -----RCARSGA--KGALLLALLLCWDIFRISAGTDCGSE-----VLPDSFSPAPAEPLFYFLQKIQDAYIVNKKPVE 66 Q9D398 -----RCARSGV--PSALLLALLLCWDITPALAVDSAGQ-----VLPDSFSPAPAEPLFYFLQKIQDAYIVNKKPVE 66 O08722 -----RCARSGA--KGALLLALLLCWDITPALAVDSAGQ-----ALPDSFSPAPAEPLFYFLQKIQDAYIVNKKPVE 66 O08747 MRKGLRATAAFCHLGLFYLLQMLVLPALALLASSTGCAQDDDFHHELETFPNDPPEPLHFLIEEAYIVNKKPVN 80 O95185 MRKGLRATAAFCHLGLFYLLQMLVLPALALLASSTGCAQDDDFHHELETFPNDPPEPLHFLIEEAYIVNKKPVN 80
25	NOV1a -----HCRAFPATQIYFKONSEWVSQNHVTVQELDEATGLRVREVQIEVSRQQVEELFGLDYNWCQVAVSSAGTTKSRRAYV 146 NOV1b -----HCRAFPATQIYFKONSEWVSQNHVTVQELDEATGLRVREVQIEVSRQQVEELFGLDYNWCQVAVSSAGTTKSRRAYV 146 Q9D398 -----HCRAFPATQIYFKONSEWVSQNHVTVQELDEATGLRVREVQIEVSRQQVEELFGLDYNWCQVAVSSAGTTKSRRAYV 146 O08722 -----HCRAFPATQIYFKONSEWVSQNHVTVQELDEATGLRVREVQIEVSRQQVEELFGLDYNWCQVAVSSAGTTKSRRAYV 146 O08747 LYCKASPATQIYFKONSEWVHKDHYVDHVDVDETSGLIVREVSIETSRQQVEELFGLDYNWCQVAVSSAGTTKSRKAYV 160 O95185 LYCKASPATQIYFKONSEWVHKDHYVDHVDVDETSGLIVREVSIETSRQQVEELFGLDYNWCQVAVSSAGTTKSRKAYV 160
30	NOV1a -----RIAYLRNFDQEPPLGKEVPLDHEVLLQCRPPGEGVPAEVEWLNEDVIDETDNTNELLTIDHNLIIQARLSDTANTTCV 226 NOV1b -----RIAYLRNFDQEPPLGKEVPLDHEVLLQCRPPGEGVPAEVEWLNEDVIDETDNTNELLTIDHNLIIQARLSDTANTTCV 226 Q9D398 -----RIAYLRNFDQEPPLGKEVPLDHEVLLQCRPPGEGVPAEVEWLNEDVIDETDNTNELLTIDHNLIIQARLSDTANTTCV 226 O08722 -----RIAYLRNFDQEPPLGKEVPLDHEVLLQCRPPGEGVPAEVEWLNEDVIDETDNTNELLTIDHNLIIQARLSDTANTTCV 226 O08747 RIAYLRNFDQEPPLGKEVSLQEVLLQCRPPGEGVPAEVEWLNEDVIDETDNTNELLTIDHNLIIQARLSDTANTTCV 240 O95185 RIAYLRNFDQEPPLGKEVSLQEVLLQCRPPGEGVPAEVEWLNEDVIDETDNTNELLTIDHNLIIQARLSDTANTTCV 240
35	NOV1a -----AHNIVAKRRSTTATVIVYVNGWSSWAENSPCSNRCGRGWQKRTCTNPAPINGGARCEGQAEUKTACTTILFVDDGAWT 306 NOV1b -----AHNIVAKRRSTTATVIVYVNGWSSWAENSPCSNRCGRGWQKRTCTNPAPINGGARCEGQAEUKTACTTILFVDDGAWT 306 Q9D398 -----AHNIVAKRRSTTATVIVYVNGWSSWAENSPCSNRCGRGWQKRTCTNPAPINGGARCEGQAEUKTACTTILFVDDGAWT 306 O08722 -----AHNIVAKRRSTTATVIVYVNGWSSWAENSPCSNRCGRGWQKRTCTNPAPINGGARCEGQAEUKTACTTILFVDDGAWT 306 O08747 AHNIVAKRRSTTATVIVYVNGWSSWAENSPCSNRCGRGWQKRTCTNPAPINGGARCEGQAEUKTACTTILFVDDGAWT 320 O95185 AHNIVAKRRSTTATVIVYVNGWSSWAENSPCSNRCGRGWQKRTCTNPAPINGGARCEGQAEUKTACTTILFVDDGAWT 320
40	NOV1a -----EWSKWSACSTECAHWRSRECHAPPPNGGRDCSGTLLDSNCTDGLCM-----LHANGLAALYAGLVVAIFV 374 NOV1b -----EWSKWSACSTECAHWRSRECHAPPPNGGRDCSGTLLDSNCTDGLCM-----LHANGLAALYAGLVVAIFV 374 Q9D398 -----EWSKWSACSTECAHWRSRECHAPPPNGGRDCSGTLLDSNCTDGLCVLNQRTLNPKSRHLETSGLVYAGLVVAIFV 386 O08722 -----EWSKWSACSTECAHWRSRECHAPPPNGGRDCSGTLLDSNCTDGLCVLNQRTLNPKSRHLETSGLVYAGLVVAIFV 386 O08747 EWSKWSACSTECATHWPRRECTAIAKNGCKKQDGLVLSKNCTDGLCM-----APDIDVVAIVVETILVIT 389 O95185 EWSKWSACSTECATHWPRRECTAIAKNGCKKQDGLVLSKNCTDGLCM-----APDIDVVAIVVETILVIT 389
45	NOV1a -----VAILMAVGVIYVRNCRDFTDITDSSAALTGGFHPVNFKTARPSNPOLLHPSVPPDLTASAGIYRGPVYALQDITDKI 454 NOV1b -----VAILMAVGVIYVRNCRDFTDITDSSAALTGGFHPVNFKTARPSNPOLLHPSVPPDLTASAGIYRGPVYALQDITDKI 454 Q9D398 -----VAILMAVGVIYVRNCRDFTDITDSSAALTGGFHPVNFKTARPSNPOLLHPSVPPDLTASAGIYRGPVYALQDITDKI 466 O08722 -----VAILMAVGVIYVRNCRDFTDITDSSAALTGGFHPVNFKTARPSNPOLLHPSVPPDLTASAGIYRGPVYALQDITDKI 466 O08747 CLATVYVLAIVYKKNHRDPESDILDS-SAINVGHCPVNTAAK-C-----LAVPPDLTSAAMPGPVYALHVSQKI 463 O95185 CLATVYVLAIVYKKNHRDPESDILDS-SAINVGHCPVNTAAK-C-----LAVPPDLTSAAMPGPVYALHVSQKI 463
50	NOV1a -----PNTNSPLLDPLSLKIKVYSSSTTSSCGPLADGALLGVLPCTYPSDFARDTHFLHLSASLGSQOLLGLPRHESVS 534 NOV1b -----PNTNSPLLDPLSLKIKVYSSSTTSSCGPLADGALLGVLPCTYPSDFARDTHFLHLSASLGSQOLLGLPRHESVS 534 Q9D398 -----PNTNSPLLDPLSLKIKVYSSSTTSSCGPLADGALLGVLPCTYPSDFARDTHFLHLSASLGSQOLLGLPRHESVS 546 O08722 -----PNTNSPLLDPLSLKIKVYSSSTTSSCGPLADGALLGVLPCTYPSDFARDTHFLHLSASLGSQOLLGLPRHESVS 546 O08747 PNTNSPLLDPLSLKIKVYSSSTTSSCGPLADGALLGVLPCTYPSDFARDTHFLHLSASLGSQOLLGLPRHESVS 533 O95185 PNTNSPLLDPLSLKIKVYSSSTTSSCGPLADGALLGVLPCTYPSDFARDTHFLHLSASLGSQOLLGLPRHESVS 533
55	NOV1a -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 614 NOV1b -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 614 Q9D398 -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 626 O08722 -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 626 O08747 PTNSNGGHLINSGVSLIPACAGIPQGVRYMYVTVYRKNRMPADSDYLLIVSGGYIPALIDKPVILVPHCA 613 O95185 PTNSNGGHLINSGVSLIPACAGIPQGVRYMYVTVYRKNRMPADSDYLLIVSGGYIPALIDKPVILVPHCA 613
60	NOV1a -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 614 NOV1b -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 614 Q9D398 -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 626 O08722 -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 626 O08747 PTNSNGGHLINSGVSLIPACAGIPQGVRYMYVTVYRKNRMPADSDYLLIVSGGYIPALIDKPVILVPHCA 613 O95185 PTNSNGGHLINSGVSLIPACAGIPQGVRYMYVTVYRKNRMPADSDYLLIVSGGYIPALIDKPVILVPHCA 613
65	NOV1a -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 614 NOV1b -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 614 Q9D398 -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 626 O08722 -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 626 O08747 PTNSNGGHLINSGVSLIPACAGIPQGVRYMYVTVYRKNRMPADSDYLLIVSGGYIPALIDKPVILVPHCA 613 O95185 PTNSNGGHLINSGVSLIPACAGIPQGVRYMYVTVYRKNRMPADSDYLLIVSGGYIPALIDKPVILVPHCA 613
70	NOV1a -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 614 NOV1b -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 614 Q9D398 -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 626 O08722 -----PTFGCLGGRISIPCTGVSLLVINGAIPQGNFYMYLLINAEESTLPLSEGTIVLSPSVTCGPTGLLLCRFVILTPHCA 626 O08747 PTNSNGGHLINSGVSLIPACAGIPQGVRYMYVTVYRKNRMPADSDYLLIVSGGYIPALIDKPVILVPHCA 613 O95185 PTNSNGGHLINSGVSLIPACAGIPQGVRYMYVTVYRKNRMPADSDYLLIVSGGYIPALIDKPVILVPHCA 613

30

40

45

50

17

||||| ||| | | ||||| +||| + ||| | + |
 Sbjct: 61 LSPVVECGPHGALFLRPVILEVPHCASLRPRDWEIVLLRSENGG 104

Table 1H. Domain Analysis of NOV1a

gnl|Pfam|pfam00791, ZU5, ZU5 domain. Domain present in ZO-1 and Unc5-like netrin receptors Domain of unknown function. (SEQ ID NO:86)
 CD-Length = 104 residues, 100.0% aligned
 Score = 147 bits (371), Expect = 3e-36

5 Query: 529 PGSSVSGTFCGLGGRLSIPGTGVSLVLPNGAIPQGKFYEMYLKAEESTLPLSEGTQTV 588
 | ||||| |||| | ||| |++| ||||| | ||++| || | | +|+
 Sbjct: 1 SGFLVSGTDFDARGGRLRGPRGTGVRLIIPGAIPQGTRYTCYLVVHDKLSTPPPLEEGETL 60

10 Query: 589 LSPSVTCGPTGLLLCRPVILTMPHCAEVSARDWIFQLKTQAHQG 632
 ||| ||| | | ||||| +||| + ||| | + |
 Sbjct: 61 LSPVVECGPHGALFLRPVILEVPHCASLRPRDWELVLLRSENGG 104

Table 1I. Domain Analysis of NOV1a

gnl|Smart|smart00005, DEATH, DEATH domain, found in proteins involved in cell death (apoptosis).; Alpha-helical domain present in a variety of proteins with apoptotic functions. Some (but not all) of these domains form homotypic and heterotypic dimers. (SEQ ID NO:87)
 CD-Length = 96 residues, 99.0% aligned
 Score = 64.7 bits (156), Expect = 2e-11

15 Query: 840 GPYAFKIPLSIRQKICNSLDAPNSRGNDWRMLAQKLSM-DRYLNIFATKAS-----PTGV 893
 | | + |++ | | + |++| ||+| + + ++ |++ +
 Sbjct: 1 PPGAASLTETREKLAKLLD--HDLGDDWRELARKLGLSEADIDQIETESPRDLAEQSYQ 58

20 Query: 894 ILDLWEALQQDDGDNLASALEEMGKSEMLVAVATD 930
 +| ||| + + | +| || +| + + + ++
 Sbjct: 59 LLRLWEQREGKNATLGTLLLEALRKMGRDDAVELLRSE 95

Table 1J. Domain Analysis of NOV1a

gnl|Smart|smart00209, TSP1, Thrombospondin type 1 repeats; Type 1 repeats in thrombospondin-1 bind and activate TGF-beta. (SEQ ID NO:88)
 CD-Length = 51 residues, 100.0% aligned
 Score = 62.4 bits (150), Expect = 1e-10

25 Query: 249 WSSWAEWSPCSNRCGRGWQKRTRCTNPAPLNGGAFCEGQAFQKTACTT-ICP 300
 | | +| ||||| || | ||| | | ||| | | + || ||
 Sbjct: 1 WGEWSEWSPCSVTCGGGVQTRTRCCNPPP--NGGGPCTGPDTETRACNEQPCP 51

Table 1K. Domain Analysis of NOV1a

gnl|Smart|smart00209, TSP1, Thrombospondin type 1 repeats; Type 1 repeats in thrombospondin-1 bind and activate TGF-beta. (SEQ ID NO:88)

CD-Length = 51 residues, 98.0% aligned
Score = 49.3 bits (116), Expect = 1e-06

Query: 305 WTEWSKWSACSTECAL-WRSRECMAPPPQNGGRDCSGTLLDSKNCTDGLC 353
| |||+|| || | ++| || ||| |+| +++ | + |
Sbjct: 1 WGEWSEWSPCSVTCGGGVQTRTRCCNPPNGGGPCTGPDTEETRACNEQPC 50

5

Table 1L. Domain Analysis of NOV1a

gnl|Pfam|pfam00531, death, Death domain. (SEQ ID NO:89)

CD-Length = 83 residues, 98.8% aligned
Score = 57.4 bits (137), Expect = 4e-09

Query: 852 QKICNSLDAPNSRGNDWRMLAQKLSM-DRYLNYPATKA----SPTGVILDLWEALQQDDG 906
+++| || | ||| ||+| + + ++ + ||| +|||| +
Sbjct: 1 RELCKLLDDP--LGRDWRRLARKLGLSEEEIDQIEHENPRLASPTYQLLDLWEQRGKNA 58
Query: 907 DLNSLASALEEMGKSEMLVAVATD 930
+ +| || +||+ + + + +
Sbjct: 59 TVGTLLEALRKMGRRDDAVELESA 82

10

Table 1M. Domain Analysis of NOV1a

gnl|Pfam|pfam00090, tsp_1, Thrombospondin type 1 domain. (SEQ ID NO:90)

CD-Length = 48 residues, 91.7% aligned
Score = 49.7 bits (117), Expect = 7e-07

Query: 250 SSWAEWSPCSNRCGRGWQKRTRTCTNPAPLNGGAFCEGQAFQKTACT 296
| |+||||| ||+| + | ||| +|| || | | + ||
Sbjct: 1 SPWSEWSPCSVTCGKGIRTRQRTCNSPA---GGKPCTGDAQETEACM 44

15

Table 1N. Domain Analysis of NOV1a

gnl|Smart|smart00409, IG, Immunoglobulin (SEQ ID NO:91)

CD-Length = 86 residues, 100.0% aligned
Score = 48.9 bits (115), Expect = 1e-06

Query: 159 PLGKEVPLDHEVLLQCRPPEGVPVAEVEWLKNEVDIDPTQDTNELLTIDHN---LIIRQA 215
| | | | | | | | | + + | ++ | |
Sbjct: 1 PPSVTVKEGESVTLSCEAS-GNPPPTVTWYKQGGKL-LAESGRFSVSRSGNSTLTISNV 58
Query: 216 RLSDTANYTCVAKNIVAKRRSTTATVIVY 244
|+ ||| | | | | |+|
Sbjct: 59 TPEDSGTYTCAATNSSGSASSGT-TLTVL 86

20

25

Table 1O. Domain Analysis of NOV1a

gnl|Smart|smart00408, IgC2, Immunoglobulin C-2 Type (SEQ ID NO:92)
 CD-Length = 63 residues, 87.3% aligned
 Score = 42.7 bits (99), Expect = 9e-05

Query: 170 VLLQCRPPEGVPVAEVEWLKNEVDIPTQDTNFLTIDHNLIRQARLSDTANYTCVAKN 229
 | | | | | | | + | | | + + ++ ++ | | + | | + | | | | + |
 Sbjct: 6 VTLTC-PASGDPVFNITWLKDGKPLPESR----VVASGSTLTIKNVSLEDSGLYTCVARN 60

5 Migration of neurons from proliferative zones to their functional sites is fundamental to the normal development of the central nervous system. Disruption of the mouse rostral cerebellar malformation mutation (rcm) gene, also called the Unc5h3 gene, resulted in a failure of tangentially migrating granule cells to recognize the rostral boundary of the cerebellum. In rcm-mutant mice, the cerebellum is smaller and has fewer folia than in
 10 wildtype, ectopic cerebellar cells are present in midbrain regions by 3 days after birth, and there are abnormalities in postnatal cerebellar-neuronal migration. Ackerman et al. (1997). Sequence analysis has revealed that the predicted rcm mouse protein is a transmembrane protein that contains 2 immunoglobulin (Ig)-like domains and 2 type I thrombospondin (THBS1) motifs in the extracellular region. Ig and THBS1 domains are also found in the
 15 extracellular region of the *C. elegans* UNC5 transmembrane protein, and the C-terminal 865-amino acid region of Rcm is 30% identical to UNC5. In addition, the UNC5 protein is essential for dorsal guidance of pioneer axons and for the movement of cells away from the netrin ligand. Ackerman et al. (1997). In the developing brain of vertebrates, netrin-1 plays a
 20 role in both cell migration and axonal guidance.

In the developing nervous system, migrating cells and axons are guided to their targets by cues in the extracellular environment. The netrins are a family of phylogenetically conserved guidance cues that can function as diffusible attractants and repellents for different classes of cells and axons. In vertebrates, insects and nematodes, members of the DCC
 25 subfamily of the immunoglobulin superfamily have been implicated as receptors that are involved in migration towards netrin sources. In *Caenorhabditis elegans*, the transmembrane protein UNC-5 has been implicated in these responses, as loss of UNC-5 function causes migration defects and ectopic expression of UNC-5 in some neurons can redirect their axons away from a netrin source.

30 The disclosed NOV1 nucleic acid of the invention encoding a UNC5H2-like protein includes the nucleic acid whose sequence is provided in Table 1A, 1C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 1A or 1C while still encoding a protein

that maintains its UNC5H2 like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes
5 nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense
10 binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 16 percent of the bases may be so changed.

The disclosed NOV1 protein of the invention includes the UNC5H2-like protein whose sequence is provided in Table 1B or 1D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table
15 1B or 1D while still encoding a protein that maintains its UNC5H2-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 9 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this UNC5H2-like protein (NOV1) may function as a member of a "UNC5H2 family". Therefore, the NOV1
20 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein
25 therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV1 nucleic acids and proteins of the invention are useful in potential
30 therapeutic applications implicated in cancer including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the UNC5H2-like protein (NOV1) may be useful in gene therapy, and the UNC5H2-like protein (NOV1) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering

from cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, Lesch-Nyhan syndrome, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, cancers, and/or other pathologies and disorders. For example, a cDNA encoding the transmembrane receptor UNC5H2-like protein may be useful in transmembrane receptor UNC5H2 therapy, and the transmembrane receptor UNC5H2-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, Lesch-Nyhan syndrome, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, cancers, and other diseases, disorders and conditions of the like. Also since this gene is expressed at a measurably higher level in several cancer cell lines (including breast cancer, CNS cancer, colon cancer, gastric cancer, lung cancer, melanoma, ovarian cancer and pancreatic cancer), it may be useful in diagnosis and treatment of these cancers. The NOV1 nucleic acid encoding the UNC5H2-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV1 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV1 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies"

section below. The disclosed NOV1 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated NOV1 epitope is from about amino acids 1 to 100. In another embodiment, a NOV1 epitope is from about amino acids 200 to 300. In further embodiments, a NOV1 epitope is from about amino acids 450 to 500, from about amino acids 600 to 900, from about amino acids 950 to 1000, from about amino acids 1200 to 1300, from about amino acids 1400 to 1600, from about amino acids 1800 to 1900, from about amino acids 1950 to 2050, and from about amino acids 2200 to 2300. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV2

NOV2 includes three novel protein tyrosine phosphatase precursor-like proteins disclosed below. The disclosed sequences have been named NOV2a, NOV2b, and NOV2c.

NOV2a

A disclosed NOV2a nucleic acid of 6994 nucleotides (also referred to as SC126422078_A) encoding a receptor protein tyrosine phosphatase precursor-like protein is shown in Table 2A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 31-33 and ending with a TAA codon at nucleotides 6874-6876. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 2A. The start and stop codons are in bold letters.

Table 2A. NOV2a nucleotide sequence (SEQ ID NO:5).

TGATTCTACTGGCTGAAAAATGTAATAAAGATGGATTTTCTTATCATTTTCTTTTACTTTTATTGGGACT
 TCAGAGACACAGGTAGATGTTTCCAATGTCGTTCCCTGGTACTAGGTACGATATAACCATCTCTTCAATTTCT
 ACAACATACACCTCACCTGTTACTAGAATAGGGGCTTCTAATGAACCAGGGCCTCCAGTCTTCTAGCCGGG
 GAAAGAGTCGGATCTGCTGGGATTCTTCTGTCTTGGGAATACACCACCTAATCCAAATGGAAGGATTATATCT
 TACATTGTCAAATATAAGGAAGTTTGTCCGTGGATGCAAACAGTATATACACAAGTCAGATCAAAGCCAGAC
 AGTCTGGAAGTTCTTCTTACTAATCTTAATCCTGGAACAACATATGAAATTAAGGTAGCTGCTGAAAACAGT
 GCTGGCATTGGAGTGTTTGTAGTATCCATTTCTCTTCCAACTGCAGAAAGTGCTCCAGGAAAAGTGGTGGAT
 TTCACAGGTGAGGCTGTCCGTTTCAGCAGTAAGCTGATGTGGTATACCTCGGCAACCAAAAAAAAAATTACC
 AGCTTCAAGATTAGTGTCAAGCATAACAGAAGTGGGATAGTAGTGAAAGAAGTGCAATCAGAGTGGAGTGC
 ATTTTAAGTGCTTCCTTCCTTGCCTGCAACGAGAATAGTGAATCTTTTATGGAGTACAGCCAGCCCT
 TCTCCAACCTTGGTAGAGTTACACCTCCATCGCGTACCACACATTCATCAAGCACGTTGACACAGAATGAG
 ATCAGCTCTGTGAAAGAGCCTATCAGTTTGTAGTGACACACTTGAGACCTTATACAACATATCTTTTGAA
 GTTTTCACTGCTACAACTGAAGCAGGTTATATTGATAGTACGATTGTGAGAACCAGAAATCAGTGCCTGAA
 GGACCACCACAAAACCTGCGTAACAGGCAACATCACAGGAAAGTCCCTTTTCAATTTATGGGACCCCAACT
 ATAGTAACAGGGAAATTTAGTTATAGAGTTGAATTATATGGACCATCAGCAGGTGCGATTTTGGATAACAGC
 ACAAAGACCTCAAGTTTGCATTCCTAACCTAACACCATTTACAATGTATGATGTCTATATTGCGGCTGAA
 ACCAGTGCAGGGACTGGGCCCAGTCAAATATTTTCAGTATTTCACTCCACCAGATGTTCCAGGGGCGAGTGT
 GATTTACAACCTGTCAGAGGTAGAATCCACGCAAGTAAGAATTACTTGGAAGAAACCACGACAACCAATGGA
 ATTATTAACCAATACCGAGTGAAGTGCTAGTTCCAGAGACAGGAATAATTTTGGAAAATACTTTGCTCACT

GGAAATAATGAGATAAATGACCCCATGGCTCCAGAAATTGTGAACATAGTACAGCCAATGGTAGGATTATAT
GAGGGTTTCAGCAGAGATGTCGTCTGACCTTCACTCACTTGCTACATTTATATATAACAGCCATCCAGATAAA
AACTTTCCGCAAGGAATAGAGCTGAAGACCAGACTTCACCAAGTTGTAACATAAGGAATCAGTATATTACT
GACATTGCAGCTGAACAGCTGACTTATGTTCTTATCAGATTAAGGAGATTTTGGGCTGAGACAATGGGGTTT
TCTAGATATACAATCATGTCATCTGCAAGCAGGGACAATTTGACTTCCCCAGGCCCTTGTGACGCCAAAAT
TTCAGAGTTACACATGTTACCATAACAGAAGTATTTTACTGCGGATCCTCCAGATCCTGTATTTTTCAT
CATTACCTTATCACTATTTTGGATGTTGAAAACCAATCCAAGAGTATTTTAAAGGACATTAAACAGTTTG
TCTCTTGTCTTATAGGGTTAAAGAAATACACAAAATACAAAATGAGAGTGGCAGCCTCAACCCACGTTGGA
GAAAGTTCTTGTCTGAAGAAAATGACATCTTTGTGAGAACTTCAGAAGATGAACCGGAATCATCACCTCAA
GATGTCGAAGTAATGATGTTACCGCAGATGAAATAAGGTTGAAGTGGTCACCACCCGAAAAGCCCAATGGG
ATCATTATTGCTTATGAAGTGCTATATAAAAAATATAGATACTTTATATATGAAGAACACATCAACAACAGAC
ATAATATTAAGGAACCTTAAGACCTCACCCCTCTATAACATTTCTGTAAGGCTTACACCAGATTTGGTGAT
GGCAATCAGGTATCTTCTTTACTCTCTGTAAGGACTTCGGAGTCAGTGCTGATAGTGCACCAGAAAATATC
ACTTACAAAATATTTCTTCTGGAGAGATTGAGCTATCATTCCTTCCCCAAGTAGTCCCAATGGAATCATA
CAAAAATATACAAATTTATCTCAAGAGAAGTAAATGGAATGAGGAAAGAACTATAAATACAACCTCTTTAAC
CAAAACATTAAAGGCTGAAGAAATATACCAATATATCATTGAGGTGTCTGCTAGTACACTCAAAGGTGAA
GGAGTTCCGAGTGTCTCCATAAGTATCTGACGGAGGAAGATGCTCTGATTCTCCCCCTCAAGACTTCTCT
GTAAAACAGTTGCTGGTGTACGGTGAAGTTGTGTCATGGCAACCACCCCTGGAGCCAATGGAATTATCCTT
TATTACACAGTTTATGCTGGAGATCATCATTAATAACTATTAATGTCACTGAAACATCATTTGGAGTTATCA
GATTTGGATTATAATGTTGAATACAGTGCTTATGTAACAGCTAGCACCAGATTGGTGATGGGAAACAAGA
AGCAATTGGAATGGGTAATAAAGCAGTGACATCATTTGAAGTATACACAGATCAAGACGTACCTGAAGGTTT
AGTTCTTTCATCAATAATTCTTTCTGGACACCTCTTCAAAACCTAATGGGATTATACAATATTACTCTGTT
TATTACAGAAATACTTCAGGTACTTTTATGAGAATTTTACTCCATGAAGTAACCAATGACTTTGACAAT
ATGACTGTATCCACAATTATAGATAAACTGACAATATTGAGTACTATACATTTTGGTTAACAGCAAGTACT
TCAGTTGGAAATGGGTAATAAAGCAGTGACATCATTTGAAGTATACACAGATCAAGACGTACCTGAAGGTTT
GTTGGAACCTGACTTACGAATCCATTTCTGCACTGCAATAAATGTAAGCTGGGTCACCCGGCTCAACCA
AACGGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGCAGACTCCTCGCCATGTGAGACCACCTCTTGTT
ACATATGAGAGAAGCATATATTTTGATAATCTGGAATAATACACTGATTATATATTAATAAATTACTCCATCA
ACAGAAAAGGATTTCTGTATACCTATACCTGAGCTATACATCAAGACTGAAGAAGATATCCAGAAACT
TCACCAATATCAACACTTTTAAAAACCTTTCTCTACCTCAGTTCTCTTATCATGGGATCCCCAGTAAAG
CCAAATGGTGAATAAATAGTTATGATTTAACTTTACAAGGACCAAAATGAAAATTTATCTTTTACTTCT
GATAATTACATAATATTTGGAAGAGCTTTCACCATTTACATTATATAGCTTTTTTGCTGCCGCAAGACTAGA
AAAGGACTTTGGTCTTCCAGTATTCTTTCTTTTACACAGATGAGTCAGTGCCGTTAGCACCTCAACAAAT
TTGACTTTAACTCACTGTACTTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGGTATT
GTTAAAGTATATAGTTTAAATTCATGAACATGAACTGACACTATATATTATAAGAAATATATCAGGATTT
AAAATGAGCCCAACTTGTGGAGTGAACAGTCAAGCCTACTCTATCCGTGTATCTGCGTTACACAAA
GTTGGAATGGCAATCAATTTAGTAATGTAGTAAATTCACAACCAAGAAATCAGTTCCAGATGTCGTGCAG
AATATGCAGTGCATGGCAACTAGCTGCGAGTCAGTTTGTAGTGAATGGGATCCACCCAAAAGGCAATGGGA
ATAATAACGCAGTATATGGTAACAGTTGAAGGAATCTACAAAAGTTCTCCCCAAGATCACATGTACACT
TTCATAAAGCTTCTTCCCAATACCTCATATGCTTTAAAGTAAGAGCTTCAACCTCAGCTGGTGAAGGTGAT
GAAAGCACATGCCATGTGAGCACACTACCTGAAACAGTTCCAGTGTTCCACAAATATTGCTTTTCTGAT
GTTACCTGAGTGCATGGCAACTTGACATGGATAAGACCTGACACTATCCTTGGCTACTTTCAAAATTACAAA
ATTACCACCTCAACTTCGTGCTCAAAAATGCAAGAATGGGAATCCGAAGAATGTGTTGAATATCAAAAAT
CAATACCTCTATGAAGCTCACTTAAGTGAAGAGACAGTATATGGATTAAAGAAATTTAGATGGTATAGATT
CAAGTGGCTGCCAGCACAATGCTGGCTATGGCAATGCTTCAAACTGGATTCTACAAAACCTCTGCTGGC
CCTCCAGATGGTCTCTGAAAATGTTGATGTAGTAGCAACATCACCTTTTAGCATCAGCATAAGCTGGAGT
GAACCTGCTGTCATTCTGGACCAACATGTTATCTGATTGATGTCAAATCGGTAGATAATGATGAATTTAAT
ATATCCTTCATCAAGTCAAATGAAGAAAATAAAACCATAGAAATTAAGATTTAGAAATATTCACAAGGTAT
TCTGTAGTGATCACTGCATTTACTGGGAACATTAGTGCTGCATATGTAGAAGGGAAGTCAAGTGTGAAATG
ATTGTTACTACTTTAGAATCAGCCCCAAGGACCCACCTAACACATGACATTTCAGAAGATACCAGATGAA
GTTACAAAATTTCAATTAACGTTCTTCTCTCTCTCAACCTAATGGAAATATCCAAGTATATCAAGCTCTG
GTTTACCGAGAAGATGATCCTACTGCTGTCCAGATTCACAACCTCAGTATTATACAGAAAACCAACACATTC
GTCATTGCAATGCTAGAAGGACTAAAAGGTGGACATACATACATATCAGTGTTTACGCAGTCAATAGTGCT
GGTGCAAGTCCAAAGGTTCCGATGAGAATAACCATGGATATCAAAGCTCCAGCAGCAACCAAAACCAACCA
ACCCCTATTATTGATGATGCCACAGGAAAACCTGCTGTGACTTCAACAACATTAACAATCAGAATGCCAATATGT
TACTACAGTGTATGATCATGGACCAATAAAAATGTACAAGTCTTGTGACAGAAACAGGAGCTCAGCATGAT
GGAAATGTAACAAAGTGGTATGATGCATATTTAATAAAGCAAGGCCATATTTTACAAATGAAGGCTTTCTCT
AACCTTCCATGTACAGAAGGAAAGACAAGGTTTGTGGCAATGAAGAAATCTACATCATAGGTGCTGATAAT
GCATGCATGATTTCTGGCAATGAAGACAAAATTTGCAATGGACCACTGAAACCAAAAAGCAATACTTATTT
AAATTTAGAGCTACAAATATTTATGGGCAATTTACTGACTCTGATTATTTCTGACCCCTGTTAAGACTTTAGGC
GAAGGACTTTGAGAAAGAACCGTAGAGATCATTTCTTCCGTCACTTTGTGTATCCTTTCAATAATTTCTCTT
GGAACAGCTATTTTGCATTTGCAAGAATTCGACAGAGCAGAAAGAGGTGGCACATACTCTCTCAGGAT
GCAGAAATTTGACACTAAATGAAGCTGGATCAGCTCATCAGTGGCAGACCTGGAACTGAAGGACGAG
AGATTAACGGGCCAATAAGCAAGAAATCTTCTGCAACATGTTGAAGAGCTTTGCACAAACAACACCTA
AAGTTTCAAGAAGAATTTTCGAATTAACAAAATTTCTTCAAGATCTTTCTTCACTGATGCTGATCTGCCT
TGGAAATAGAGCAAAAACCGCTTCCCAACATATAAACCATATAATAAACAGAGTAAAGCTGATAGCTGAC
GCTAGTGTTCAGGTTCCGATTATATTAATGCCAGCTATATTTCTGGTTATTTATGTCCAAATGAATTTAT
GCTACTCAAGGCTCACTACCAGGAACAGTTGGAGATTTTGGAGAATGGTGTGGGAAACAGAGCAACAAACA
TTAGTAATGCTAACACAGTGTTTTGAAGAAGGACGGATCAGATGCCATCAGTATTGGCCAGAGGACAACAG

```
CCAGTTACTGTCTTTGGAGATATAGTGATTACAAAGCTAATGGAGGATGTTCAAATAGATTGGACTATCAGG
GATCTGAAATTTGAAAGGCATGGGGATTGCATGACTGTTTCGACAGTGAACCTTTACTGCCTGGCCAGAGCAT
GGGGTTCCTGAGAACAGCGCCCCCTCTAATTCACCTTTGTGAAGTTGGTTCGAGCAAGCAGGGCACATGACACC
ACACCTATGATTGTTCACTGCAGTGTCTGGAGTTGGAAGAACTGGAGTTTTTATTGCTCTGGACCATTAAACA
CAACATATAAATGACCATGATTTTGTGGATATATATGGACTAGTAGCTGAAGTGAAGAATGTGC
ATGGTGCAGAATCTGGCACAGTATATCTTTTACACCACTGCATTCTGGATCTCTATCAAATAAGGGAAGT
AATCAGCCCATCTGTTTTGTTAACTATTTCAGCACTTCAGAAGATGGACTCTTTGGACGCCATGGAAGTGGT
GATGTTGAGCTTGAATGGGAAGAAACCACTATGTAAATATTTCAGACCAAAGGATACAATTGGAAGAGATTTT
TAAATCCCAGGGGCCAAAGTTACCCCTCATTCTCCGAATTGAAATGTGCAACCTTAAAGAAATATCTATG
CTTCTCTCAC
```

In a search of public sequence databases, the NOV2a nucleic acid sequence, located on chromosome 12 has 777 of 3293 bases (84%) identical to a gb:GENBANK-
ID:AF063249|acc:AF063249.1 mRNA from *Rattus norvegicus* (*Rattus norvegicus* glomerular
mesangial cell receptor protein-tyrosine phosphatase precursor (PTPRQ) mRNA, complete
cds) (E = 0.0). Public nucleotide databases include all GenBank databases and the GeneSeq
patent database.

The disclosed NOV2a polypeptide (SEQ ID NO:6) encoded by SEQ ID NO:5 has 2281
amino acid residues and is presented in Table 2B using the one-letter amino acid code. Signal
P, Psort and/or Hydropathy results predict that NOV2a has a signal peptide and is likely to be
localized in the plasma membrane with a certainty of 0.4600. In other embodiments, NOV2a
may also be localized to the microbody (peroxisome) with a certainty of 0.1381, the
endoplasmic reticulum (membrane) with a certainty of 0.1000 or in the endoplasmic reticulum
(lumen) with a certainty of 0.1000. The most likely cleavage site for a NOV2a peptide is
between amino acids 17 and 18, at: SET-QV.

Table 2B. Encoded NOV2a protein sequence (SEQ ID NO:6).

```
MDFLIIIFLLFIQTSETQVDVSNVPGTRYDITISSISTYTSVPVTRIGASNEPFPVFLAGERVGSAGILL
SWNTFPNPNGRISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTNLNPQTYYEIKVAAENSAGIGVFSDFP
LFQTAEASAPGKVVDFTGEAVPFSSKLMYTSATKKITSFKISVKHNRSGIVVKEVSRVEICLSASPLHC
NENSESFLWSTASPSPTLGRVTPPSRTHSSSTLTQNEISSVKEPISFVVTHLRPYTTYLFEVSAATTEAGY
IDSTIVRTPESVPEGPQNCVTGNITGKSFSILWDPPTIVTGKFSYRVELYGPSAGRI LDNSTDKLKFATFN
LTPFTMYDVYIAAETSAGTGPKSNISVFTPPDVPVGAFFDLQLAEVESTQVRITWKKPRQPNGIINQYRVKVL
VPETGIIILENTLLTGNNEINDEMAPEIVNIVQPMVGLYEGSAEMSSDLHSLATFIYN SHPDKNFPARNRAED
QTSPPVVTTRNQYITDIAAEQLTYVLIRLRFFWAETMGFSRYTIMSSASRDNLSPGPLSAQNFRVTHVTITE
VFLHWDPPDPVFFHHYLITILDVENQSKSIILRLTNSLSLVLI GLKKYTKYKMRVAASTHVGESSLSEENDI
FVRTSEDEPESSPDVEVIDVTADEIRLKWSPPEKPNGIIIAVEVLYKNIDTLYMKNTSTTDIILRLNLRPHT
LYNISVRSYTRFGHGNQVSSLLSVRTSESVPDSAPENITYKNISSGEIELSFLPPSSPNGIIQKYTIYLRKS
NGNEERTINTSLTQNIKGLKKYTOYIEVSASTLKGEGVRSAPISILTEEDAPDSPQDFSVKQLSGVTVK
LSWQPPLEPNGIILYTVYVWRSSSLKTINVTETSLSDLDNVEYSAYVTASTRFGDGKTRSNIIISFQTPF
GPSDPKPKDVYANLSSSSIIIFWTPPSKPNGIIQYYSVYYRNTSGTFMQNFTLHEVTNDFDNMTVSTIIDKL
TIFYSTFWLTA SVGNKSSDIEVYTDQDVPGEFVGNLTYESISS TAINVSWVPPAQPNGLVFFYVSL
ILQQTFRHVRPPLVTFYERSIYFDNLEKYTDYILKITPSTEGFSDTYTAQLYIKTEEDIPETSPIINTFKNL
SSTS VLLSWDPFVKPNGAIIISYDLTLQGPENYSFITS DNYIIEELSFFTLYSFFAAARTKGLGPS SIF
FYTDESVP LAPPNLTLINCTSDFWLKWSPSP LPPGIVKVYSFKIHEHETDTIYKNIISGFKTEAKLVGLE
PVSTYSIRVSAFTKVGNGNQFSNVVKETTQESVPDVQNMQCMATSWQSVLVKWDPPKKANGIITQYMTVE
RNSTKVSPQDHMYTFKLLANTSYVFKVRAST SAGEGDESTCHVSTLPETVPSVPTNIAFSDVQSTATLTW
IRPDTILGYFQNYKITTLQRAQCKEWESEECVEYQKIQYLYEAHLTEETVYGLKKFRWYRFQVAASTNAGY
GNASNWISTKTLPGPPDGPENHVH VATSPFISISWSEPAVITGPTCYLIDVKSVDNDEFNISFIKSNEEN
KTIEIKDLEIFTRYSVVITAFTGNISAA YVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKFQLTFLP
PSQPNGINIQVYQALVYREDDPTAVQIHNLSIIQKNTNFVIA MLEGLKGHTYNI SVYAVNSAGAGPKVPMRI
```

```

TMDIKAPARPKTKPTPIYDATGKLLVTSTTITIRMPICYSDDHGPIKNVQLVTETGAQHDGNVTKWYDAY
FNKARPYFTNEGFFNPCTEGKTKFSGNEEIIYIIGADNACMIPGNEDKICNGPLPKPKQYLFKFRATNIMGQ
FTDSDYSDPVKTLGEGLSERTVEIILSVTLCILSIILLGTAIFAFARIROKQKEGGTYSQDAEIIIDTKLKL
DQLITVADLELKDERLTRPI SKKSFLQHVVELCTNNNLKFQEEFSELPKFLQDLSSTADLPWNRAKNRFPN
IKPYNNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLPGTVGDFWRMVWETRAKTLVMLTQCFEK
GRIRCHQYWPEDNKPVTVFGDIVITKLMEVDQIDWTIRDLKIERHGDGMTVRQCNETAWPEHGVPENSAPLI
HFVKLVRSRAHDTTPMIVHCSAGVGRGTGVFIALDHLTQHINDHDFVDIYGLVAELRSEMRMCMVQNLAQYIF
LHQCILDLLSNKGSNQPICFVNYSALQKMDSLDAMEGGDVELEWEETTM

```

A search of sequence databases reveals that the NOV2a amino acid sequence has 1894 of 2301 amino acid residues (82%) identical to, and 2078 of 2301 amino acid residues (90%) similar to, the 2302 amino acid residue ptnr:SPTREMBL-ACC:O88488 protein from *Rattus norvegicus* (Rat) (Glomerular Mesangial Cell Receptor Protein-Tyrosine Phosphatase Precursor (EC 3.1.3.48)) (E = 0.0). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV2 is expressed in at least kidney and colon. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, the sequence is predicted to be expressed in *Rattus norvegicus* :kidney because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AF063249|acc:AF063249.1) a closely related *Rattus norvegicus* glomerular mesangial cell receptor protein-tyrosine phosphatase precursor (PTPRQ) mRNA, complete cds homolog.

NOV2b

A disclosed NOV2b nucleic acid of 2565 nucleotides (also referred to as CG50718-02) encoding a novel Glomerular Mesangial Cell Receptor Protein-Tyrosine-like protein is shown in Table 2C. An open reading frame was identified beginning with an AGA codon at nucleotides 1-3 and ending with a GAG codon at nucleotides 2563-2565. The start and stop codons are in bold letters in Table 2C. Because the first and last codons are not traditional initiation and termination codons, NOV2b could represent a partial reading frame that extends in the 5' and/or 3' directions.

Table 2C. NOV2b nucleotide sequence (SEQ ID NO:7).

```

AGATCTCCTGAAGGGTTTGTGGAAACCTGACTTACGAATCCATTTTCGTCAACTGCAATAAATGTAAGCTGG
GTCCCAACCGCTCAACCAAACGGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGCAGACTCCTCGCCAT
GTGAGACCACCTCTTGTACATATGAGAGAAGCATATATTTTGATAATCTGGAAAAATACACTGATTATATA
TTAAAAATTACTCCATCAACAGAAAAGGGATTCTCTGATACCTATACTGCCAGCTATACATCAAGACTGAA
GAAGATGTCCCAAGAACTTCACCAATAATCAACACTTTTAAAAACCTTTCCTCTACCTCAGTTCTCTTATCA
TGGGATCCCCAGTAAGCCAAATGGTGCAATAATAAGTTATGATTTAACTTTACAAGGACCAATGAAAAAT
TATTTCTTTCATTACTTCTGATAATTACATAATATTGGAAGAGCTTTCACCATTTACATTATATAGCTTTTTT
GCTGCCGCAAGAACTAGAAAAGGACTTGGTCCTTCCAGTATTCTTTCTTTTACACAGATGAGTCAGTGCCG

```



```

TTAGCACCTCCACAAAATTGACTTTAATCAACTGTACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGT
CCTCTTCCAGGTGGTATTGTTAAAGTATATAGTTTAAAAATTCATGAACATGAACTGACACTATATATTAT
AAGAATATATCAGGATTTAAACCTGAAGCCAAACTTGTTGGACTGGAACCACTCAGCACCTACTCTATCCGT
GTATCTGCGTTCCACAAAGTTGGAATGGCAATCAATTTAGTAATGTAGTAAAATTCACAACCAAGAATCA
GTTCCAGATGTCGTGCAGAAATATGCAGTGCATGGCAACTAGCTGGCAGTCAGTTTTAGTGAATGGGATCCA
CCCAAAAGGCAATGGAATAATAACGCAGTATATGGTAACAGTTGAAAGGAATTCACAAAAGTTTCTCCC
CAAGATCACATGTACACTTTCATAAAGCTTCTTGCCAATACCTCATATGTCTTTAAAGTAAGAGCTTCAACC
TCAGCTGGTGAAGGTGATGAAAGCACATGCCATGTCAGCACACTACCTGAAACAGTTCCCAGTGTTCACACA
AATATTGCTTTTTCTGATGTTCACTCAACTAGTGCAACATGACATGGATAAGACCTGACACTATCCTTGGC
TACTTTCAAAATTACAAATTACCACTCAACTTCGTGCTCAAAATGCAAGAATGGGAATCCGAAGAATGT
GTTGAATATCAAAAATTCAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGGATTAAAGAAA
TTTAGATGGTATAGATTCCAAGTGGCTGCCAGCACCATGCTGGCTATGGCAATGCTTCAAAGTGGATTCT
ACAAAACTCTGCCTGGCCCTCCAGATGGTCTCTGAAATGTTTCATGTAGTAGCAACATCACCTTTTAGC
ATCAGCATAGCTGGAGTGAACCTGCTGTCACTTACTGGACCAACATGTTATCTGATTGATGTCAAATCGGTA
GATAATGATGAATTTAATATATCCTTCATCAAGTCAAATGAAGAAAATAAAACCATAGAAATTAAGATTTA
GAAATATTCACAAGGTATTCTGTAGTGATCACTGCATTTACTGGGAACATTAGTGCTGCATATGTAGAGGG
AAGTCAAGTGCTGAAATGATTGTTACTACTTTAGAATCAGCCCCAAAGGACCCACCTAACACATGACATTT
CAGAAGATACCAGATGAAGTTACAAAATTTCAATTAACGTCCCTTCTCCTTCTCAACCTAATGGAAATATC
CAAGTATATCAAGCTCTGGTTTACCGAGAAGATGATCCTACTGCTGTCCAGATTACAACCTCAGTATTATA
CAGAAAACCAACACATTCGTCATTGCAATGCTAGAAGGACTAAAAGGTGGACATACATACATATCAGTGTT
TACGCAGTCAATAGTGCTGGTGCAGGTCCAAAGGTTCCGATGAGAATAACCATGGATATCAAAGCTCCAGCA
CGAGCAAAAACCAACCAACCCCTATTATGATGCCACAGGAAAACCTGCTGTGACTCAACACAAATATACA
ATCAGAATGCCAATATGTTACTACAGTGATGATCATGGACCAATAAAAAATGTACAAGTGCTTGTGACAGAA
ACAGGAGCTCAGCATGATGAAATGTAACAAAGTGGTATGATGCATATTTAATAAAGCAAGGCCATATTTT
ACAAATGAAGGCTTTCTTACCCTCCATGTACAGAAGGAAAGACAAAGTTTGTAGTGGCAATGAAGAAATCTAC
ATCATAGTGCTGATATGCATGATGATGCTGCGCAATGAAGACAAAATTTGCAATGGACCACTGAAACCA
AAAAGCAATACTTATTTAAATTTAGAGCTACAAATATTATGGGACAATTTACTGACTCTGATTATTCTGAC
CCTGTTAAGACTTTAGGCGAAGGACTTTCAGAAAGACCCTCGAG

```

The disclosed NOV2b polypeptide (SEQ ID NO:8) encoded by SEQ ID NO:7 has 855 amino acid residues and is presented in Table 2D using the one-letter amino acid code.

Table 2D. Encoded NOV2b protein sequence (SEQ ID NO:8).

```

RSPEGEVGNLTYESISSSTAINVSWVPPAQPNGLVFFYYVSLILQQTPRHVRPPLVITYERSIYFDNLEKYTDYI
LKITPSTEKGFSDTYTAQLYIKTEEDVPESTPIINTFKNLSSTSVLLSWDPPVKPENGAIISYDLTLQGPNE
YSFITSNDNYIILEELSPFTLYSFFAAARTRKGLGPSSILFFYTDESVPPLAPPQNLTINCTSDFWLKWSPS
PLPGGIVKVSFKIHEHETDTIYYKNISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNNGNQFSNVVKFTTQES
VPDVVQNMQCMATSWQSVLVKWDPPPKKANGIITQYMTVERNSTKVSPODHMYTFIKLLANTSYVFKVRAST
SAGEGDESTCHVSTLPETVPSVPTNIAFSDVQSTSATLTWIRPDTILGYFQNYKITTLQRAQKCKEWESEEC
VEYQKIQLYLAEHLTEETVYGLKKFRWYRFQVAASTNAGYGNASNWISTKTLPGPPDGPPENVHVATSPFS
ISISWSEPAVITGPTCYLIDVKSVDNDEFNISFIKSNEENKTIKDLIEIFTRYSVVITAFNGNISAAAYVEG
KSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKFQLTSLPPSQFNGNIQVYQALVYREDPTAVQIHNLISII
QKTNTFVIAMLEGLKGGHTYNI SVYAVNSAGAGPKVPMRITMDIKAPARPKTKPTPIYDATGKLLVTSTTIT
IRMPICYSSDDHGPIKNVQVLVTETGAQHDGNVTKWYDAYFNKARPYFTNEGFPNPPCTEGKTKFSGNEEYI
IIGADNACMIPGNEDKICNGPLKPKKQYLFKFRATNIMGQFTSDSYSDPVKTLGEGLSERTLE

```

NOV2b is expressed in Brain, Colon, Fetal brain, Germ Cell, Heart, Kidney, Prostate, Uterus, brain, breast, colon, kidney, lung.

NOV2c

A disclosed NOV2c nucleic acid of 6903 nucleotides (also referred to as CG50718-05) encoding a novel Glomerular Mesangial Cell Receptor Protein-Tyrosine Phosphatase Precursor-like protein is shown in Table 2E. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 1-3 and ending with a TGA codon at nucleotides

6901-6903. A putative untranslated regions upstream from the initiation codon and downstream of the termination codon are underlined in Table 2E. The start and stop codons are in bold letters.

Table 2E. NOV2c nucleotide sequence (SEQ ID NO:9).

ATGGATTTTCTTATCATTTTCTTTTACTTTTATTGGGACTTCAGAGACACAGGTAGATGTTTCCAATGTC
 GTTCTCGTACTAGGTACGATATAACCATCTCTTCAATTTCTACAACATACACCTCACCTGTTACTAGAATA
 GTGACAACAAATGTAACGAGAACCAGGGCCCTCCAGTCTTCTAGCCGGGAAAGAGTCGGATCTGCTGGGATT
 CTTCTGTCTTGGAAACACCACTAATCCAAATGGAAGGATTATATCTTACATTGTCAAATATAAGGAAGTT
 TGTCCGTGGATGCAACAGTATATACACAAGTCAGATCAAAGCCAGACAGTCTGGAAGTTCTTCTTACTAAT
 CTTAATCCTGGAACAACATATGAAATTAAGGTAGCTGCTGAAAACAGTGTGGCATTGGAGTGTGTTAGTGAT
 CCATTTCTCTTCCAACTGCAGAAAGTCCAGCTCCAGGAAAGTGGTGAATCTCACAGTTGAGGCCCTACAAC
 GCTTCAGCAGTTAAGCTGATTTGGTATTTACCTCGGCAACCAATGGCAAAATACCAGCTTCAAGATTAGT
 GTCAGCATGCCAGAAGTGGGATAGTAGTGAAGATGTCTCAATCAGAGTAGAGGACATTTTACTGGGAAA
 TTGCCAGAATGCCAGCAAGTAAAGATTACTTGGAAAGAAACACGACAACCAATGGAATTATTACCAATAC
 AGAGTTACACCTCCATCGGTACACACATTATCAAGCAGCTTGACACAGAATGAGATCAGCTCTGTGTGG
 AAAGAGCCTATCAGTTTTGTAGTGACACACTTGAGACCTTATACAACATATCTTTTGAAGTTTCAGCTGCT
 ACAACTGAAGCAGGTTATATTGATAGTACGATTGTGAGAACACCAAGATCAGTGCCTGAAGGACCACACAA
 AACTGCGTAACAGGCAACATCACAGGAAAGTCCCTTTTCAATTTTATGGGACCCACCACTATAGTAACAGGG
 AAATTTAGTTATAGAGTTGAATTATATGGACCATCAGGTGCGATTTTGGATAACAGCACAAAAGACCTCAAG
 TTTGCATTCACTAACCTAACACCATTTACATGTATGATGTCTATATTGCGGCTGAAACCAAGTGCAGGGACT
 GGGCCCAAGTCAATATTTTCACTTCCACCATGTTCCAGGGGCAAGTGTGATTTACAACCTTGCA
 GAGGTAGATCCAGGCAAGTAAAGATTACTTGGAAAGAAACACGACAACCAATGGAATTATTACCAATAC
 CGAGTGAAAGTGTAGTTCCAGAGACAGGAATAATTTGGAAATACTTTGCTCACTGGAAATAATGAGATA
 AATGACCCCATGGGTCCAGAAATTGTGAACATAGTAGAGCAATGGTAGGATTATATGAGGGTTCCAGAGAG
 ATGTCGTCTGACCTTCACTCACTTGCTACATTTATATATAACAGCCATCCAGATAAAAACTTCTGCAAGG
 AATAGAGCTGAAGACCACTTCCAGCTTGAACCTACAAGGAATCAGTATATTACTGACATTGCAAGTGA
 CAGCTGTCTTATGTTTATCAGGAGACTTGTACCTTTCACTGAGCACATGATTAGTGTATCTGCTTTCAACATC
 ATGGGAGAAGGACCACCAACAGTTCTCAGTGTAGGACACGTGAGCAAGTGCCAAAGTCCATTAAAAATTATA
 AACTATAAAAAATATTAGTTCTTCTATCTATTTTGTATATTGGGATCCTCCAGAAATATCCCAATGGAAAAATA
 ACTCACTATACGATTTATGCAATGGAATTGGATACAAACAGAGCATTCCAGATAACTACCATAGATAACAGC
 TTTCTCATAAACAGGTATAGGGTTAAAGAAATACACAAAATACAAATGAGAGTGGCAGCCTCAACCCAGGTT
 GGAGAAAGTCTTTGTCTGAAGAAATGACATCTTTGTGAGAACTTCAGAAGATGAACCGGAATCATCACCT
 CAAGATGTCGAAGTAATTGATGTTACCGCAGATGAAATAAGGTTGAAGTGGTCACCACCCGAAAGCCCAAT
 GGGATCATATTGCTTATGAAGTGTATATAAAAAATAGATACTTTATATATGAAGAACACATCAACAACA
 GACATAATATTAAAGGAACCTTAAGACCTCACACCCCTCTATAACATTTCTGTAAGGTCTTACACAGATTGGT
 CATGGCAATCAGGTATCTTCTTACTCTCTGTAAGGACTTCGGAGACTGTGCCTGATAGTGCACCAGAAAAAT
 ATCACTTACAAAAATATTTCTTCTGAGAGATTGAGCTATCATTCTTCCCCCAAGTAGTCCCAATGGAATC
 ATACAAAAATATACAATTTATCTCAGAGAAGTAATGGAATGAGGAAAGAACTATAAATACAACCTCTTTA
 ACCCAAAACATTCTGAAGAAATATACCCAATATATCATTGAGGTGTCTGCTAGTACACTCAAAGGTGAAGGA
 GTTCGGAGTGTCTCCATAAGTATACCTGACGGAGGAAGATGCTCCTGATTCTCCCCCTCAAGACTTCTCTGTA
 AAACAGTTGTCTGGTGTACGGTGAAGTTGTCTATGGCAACCAACCCCTGGAGCCAAATGGAATTATCCTTTAT
 TACACAGTTTATGCTGGAGGAATAGATCATCATTAAAACTATTAAATGTCACTGAACATCATTTGGAGTTA
 TCAGATTGGATTATAATGTTGAATACAGTGCTTATGTAACAGCTAGCACCAGATTTGGTGATGGGAAAAACA
 AGAAGCAATATCATTAGCTTTCAAACACCAAGAGGACCAAGCGATCCTCCCAAGATGTTTATTATGCAAAC
 CTCAGTTCTTCAATAAATCTTTCTGGACACCTCCTTCAAACCTAATGGGATTATACAATATTACTCT
 GTTTATTACAGAAATACTTCAAGTACTTTTATGCAAGATTTTACACTCCATGAAGTAACCAATGACTTTGAC
 AATATGACTGTATCCACAATTATAGATAAACTGACAATATTCAGCTACTATACATTTTGGTTAACAGCAAGT
 ACTTCAGTTGGAATGGGAATAAAAGCAGTGACATCATTGAAGTATACACAGATCAAGACGTCCTGAAAGGG
 TTTGTTGGAACCTGACTTACGAATCCATTTCTGCAACTGCAATAAATGTAAGCTGGGTCCCACCGGCTCAA
 CCAAACGGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGCAGACTCCTCGCCATGTGAGACCACCTCTT
 GTTACATATGAGAGAAGCATATATTTGATAATCTGGAAAAATACACTGATTATATATTAATAATTACTCCA
 TCAACGATAAAGGGATTCTCTGATACCTATCTGCCCAGCTATACATCAAGACTGAAGAAAGATGTCCCAGAA
 ACTTCACCAATAATCAACACTTTTAAAAACCTTTCTCTACCTCAGTTCTCTTATCATGGGATCCCCCAGTA
 AAGCCAAATGGTGAATAATAAGTTATGATTTAACTTTACAAGGACCAATGAAATATTCTTTTCACTACT
 TCTGATAATTACATAATATTGGAAGAGCTTTCAACATTTACATTATATAGCTTTTGTGTCGCCGAAGAAT
 AGAAAAGGACTTGGTCTTCCAGTATCTTTTCTTACACAGATGAGTCAGTGCCGTTAGCACCTCCAGAA
 AATTTGACTTTAATCAACTGTACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGGT
 ATTGTTAAAGTATATAGTTTAAAAATTATGAACATGAACTGACATATATATATAAGAATATATCAGGA
 TTTAAACTGGAAGCCAACTTTGTTGGACTGGAACCAAGTCAGACCTACTCTATCCGTGTATCTGCGTTCAAC
 AAGTTGGAAATGGCAATCAATTTAGTAATGTAGTAAATTCACAACCAAGAAATCAGTTCCAGATGTGCTG
 CAGAAATGACAGTGATGGCAACTAGCTGGCAGTCAGTTTGTAGTGAATGGGATCCACCCAAAAGGCAAT
 GGAATAATAACGCAGTATATGGTAACAGTTGAAAGGAATCTACAAAAGTTTCTCCCAAGATCACATGTAC
 ACTTTTATAAAGCTTCTTGCCAATACCTCATATGCTTTAAAGTAAGAGCTTCAACCTCAGCTGGTGAAGT

```

GATGAAAGCACATGCCATGTCAGCACACTACCTGAAACAGTTCCAGTGTTCACCAAAATATTGCTTTTCT
GATGTTTCAGTCAACTAGTGCACATGACATGGATAAGACCTGACACTATCCTTGGCTACTTTCAAAATTAC
AAAATTACCACTCAACTTCGTGCTCAAAAATGCAAAGATGGGAATCCGAAGAATGTGTTGAATATCAAAA
ATTCAATACCTCTATGAAGCTCACTTAAGAGACAGTATATGGATTAAAGAAATTTAGATGGTATAGA
TTCCAAGTGGCTGCCAGCACCAATGCTGGCTATGGCAATGCTTCAAAGTGGATTCTACAAAACCTGCGCT
GGCCCTCCAGATGGTCTCTGAAATGTTGATGTAGTAGCAACATCACCTTTTAGCATCAGCATAAGCTGG
AGTGAACCTGCTGTCATTACTGGACCAACATGTTATCTGATTGATGTCAAATCGGTAGATAATGATGAATTT
AATATATCCTTCATCAAGTCAAATGAAGAAAATAAAACCATAGAAATTAAGATTTAGAAATATTCACAAGG
TATTCGTAGTGTGATCACTGCATTTACTGGGAACATTAGTGCTGCATATGTAGAAGGGAAGTCAAGTGTGAA
ATGATTGTTACTACTTTAGAATCAGCCCCAAAGGACCCACCTAACACATGACATTTCAGAAGATACCAGAT
GAAGTTACAAAATTTCAATTAACGTCCTTCTCTCTCTCAACCTAATGGAATATCCAAGTATATCAAGCT
CTGGTTTACCAGAGAAGATGATCCTACTGCTGTCCAGATTCAACACCTCAGTATTATACAGAAAACCAACACA
TTGTCATTGTCATGCTAGAGGACTAAAAGGTGGACATACATACATATCAGTGTTTACGCAGTCAATAGT
GCTGGTGCAGGTCCAAAGGTTCCGATGAGAATAACCATGGATATCAAAGCTCCAGCAGCAGCAAAAACCAAA
CCAACCCCTATTTATGATGCCACAGGAAACTGCTTGTGACTTCAACAACAATTACAAATCAGAAATGCCAATA
TGTTACTACAGTGTGATCATGGACCAATAAAAAATGTACAAGTGTCTGTGACAGAAACAGGAGCTCAGCAT
GATGGAATGTAACAAAGTGGTATGATGATATTTAATAAAGCAAGGCCATATTTTACAAATGAAGGCTTT
CCTAACCCCTCCATGTACAGAAGGAAAGACAAAGTTTAGTGGCAATGAAGAAATCTACATCATAGGTGCTGAT
AATGCATGCATGATTCCTGGCAATGAAGACAAAATTTGCAATGGACCACTGAAACCAAAAAAGCAATACTTA
TTTAAATTTAGAGCTACAAATATTATGGGACAATTTACTGACTCTGATTATTCTGACCTGTAAAGACTTTA
GGCGAAGGACTTTAGAAAGAACCTTAGAGATCATTCTTTCCGTCACTTTGTGTATCCTTTCAATAATTTCTC
CTTGGAACACAGCTATTTTTCGATTTGCAAGAAATCGACAGAAGCAGAAGAAGGTGGCACATACTCTCTCAG
GATGCAGAAATATTGACACTAAATGAAGCTGGATCAGCTCATCAGTGGCAGACCTGGAATGAAGGAC
GAGAGATTACGCGGTACTTAGTTATAGAAAATCCATCAAGCCAAATAAGCAAGAAATCCTTCTGCAACAT
GTTGAAGAGCTTTGCACAAACAACCTAAGTTTCAAGAAGAAATTTTCGGAATTACCAAAATTTCTTCAG
GATCTTTCTTCAACTGATGCTGATCTGCCTTGGAAATAGAGCAAAAAACCGCTTCCCAACATAAAACCATAT
AATAATAACAGAGTAAAGCTGATAGCTGACGCTAGTGTCCAGGTTCCGATTATATTAAATGCCAGCTATATT
TCTGTTATTTATGTCCAAATGAATTTATTGCTACTCAAGTCCACTACCAGGAACAGTTGGAGATTTTGG
AGAATGGTGTGGGAACAGAGCAAAACATTAGTAATGCTAACACAGTGTTTTGA AAAAGGACGGATCAGA
TGCCATCAGTATTGGCCAGAGGACAACAAGCCAGTTACTGTCTTTGGAGATATAGTGATTACAAAGCTAATG
GAGGATGTTCAAATAGATTGGACTATCAGGGATCTGAAAATGAAAGGCATGGGGATTGCATGACTGTTTGA
CAGTGTAACCTTACTGCCTGGCCAGAGCATGGGGTTCTGAGAACAGCGCCCTCTAATTCACCTTTGTGAAG
TTGGTTCGAGCAAGCAGGGCACATGACACCACACCTATGATTGTTCACTGTAGTGTGGAGTTGGAAGAACT
GGAGTTTTTATTGCTCTGGACATTTAACACAACATATAAATGACCATGATTTTGTGGATATATGGACTA
GTAGCTGAACCTGAGAAGTGAAGAATGTGCATGGTGCAGAACTGGGCACAGTATATCTTTTACACCAGTGC
ATTCTGGATCTCTTATCAAATAAGGGAAGTAATCAGCCCATCTGTTTGTTAACATATTCAGCACTTCAGAAG
ATGGACTCTTTGGACCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGAAACCACTATGTAA

```

In a search of public sequence databases, the NOV2c nucleic acid sequence, located on chromosome 12 has 5903 of 6906 bases (85%) identical to a gb:GENBANK-

ID:AF063249|acc:AF063249.1 mRNA from *Rattus norvegicus* (*Rattus norvegicus* glomerular mesangial cell receptor protein-tyrosine phosphatase precursor (PTPRQ) mRNA, complete cds) (E=0.0). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV2c polypeptide (SEQ ID NO:10) encoded by SEQ ID NO:9 has 2300 amino acid residues and is presented in Table 2F using the one-letter amino acid code.

Signal P, Psort and/or Hydropathy results predict that NOV2c has a signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.4600. In other embodiments, NOV2c may also be localized to the microbody (peroxisome) with a certainty of 0.1260, the endoplasmic reticulum (membrane) with a certainty of 0.1000 or in the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for a NOV2c peptide is between amino acids 17 and 18, at: SET-QV.

Table 2F. Encoded NOV2c protein sequence (SEQ ID NO:10).

```

MDFLIIFLLFFIGTSETQVDVSNVVPGRYDITISSISTTYTSPVTRIVTNTVTEPGPPVFLAGERVGSAGI
LLSWNTPPNPNGRIISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTNLNPGETTYEIKVAAENSAGIGVFS
PFLFQTAEAPAGKVNNLTVEAYNASAVKLIWYLPQPNKGKITSFKISVKHARSGIVVKDVSIRVEDILTGK
LPECNVENSESFLWSTASPSPTLGRVTPPSRTTHSSSTLTQNEISSVWKEPISFVTHLRPYTTYLFEEVSAA
TTEAGYIDSTIVRTPESVPEGPPQNCVGTGNITGKSFSILWDPPTIVTGKFSYRVELYGPSGRILDNSTKDLK
FAFTNLTPFTMYDVYIAAETSAGTGPKSNISVFTPPDVGAVFDLQLAEVESTQVRITWKKPRQPNGINQY
RVKVLVPETGIILENTLLTGNNEINDPMAPEIVNIVEPMVGLYEGSAEMSSDLHSLATFIYNSHPDKNFPAR
NRAEDQTSFVVTTNRNQYITDIAEQLSYVIRRLVFTEHMISVSAFTIMGEGPPTVLSVRTRQQVPSSIKII
NYKNISSSSILLYWDPPEYPNGKITHYTIYAMELDTNRAFQITTTIDNSFLITIGIGLKKYTKYKMRVAASTHV
GESSLSEENDIFVTSDEDEPSSPDVEVIDVTADEIRLKWSPPEKPNGIIAYEVLYKNIDTLYMKNSTST
DIILRNLRPHLTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGEIELSFLPPSSPNGI
IQKTYTIYLRKSNNGNEERTINTTSLTONILKKYQYIIIEVASTLKGEVRSAPISILTEEDAPDSPPQDFSV
KQLSGVTVKLSWIRPDPLENGIILYTYVVRNRSSLKTINVTETSLELSDLDYNVEYSAYVTASTRFGDGKT
RSNIISFQTPEGSPDPKDVYANLSSSIIILFWTPPSKNGIIQYYSVYRNTSGTFMONEFTLHEVTNDFD
NMTVSTIIDKLTIFFSYTFFWLTASTSVGNGNKSDDIEVYTDQDVEGFGVNLTYESISSSTAINVSWVPPAQ
PNGLVFFYYVSLILQQTPRHVRPPLVTYERSIYFDNLEKYTDYILKITPSTKEGFSPTYTAQLYIKTEEDVPE
TSPINTFKNLSSTSVLLSWDPPVKPNGAIIISYDLTLQGNENYSFITSNDYIILEELSPFTLYSFFAAART
RKGLGPSSILFFYTDESVPPLAPPQNLTLINCTSDFWLKWSPSPPLGGIVKVYSFKIHEHETDTIYKNISG
FKTEAKLVGLEPVSTYSIRVSAFTKVGNQFSNVVKKFTQESVPDVVQNMQCMATSWQSVLVKWDPPKKAN
GIITQYMTVERNSTKVPQDHMYTFIKLLANTSYVFKVRASTSAGEDESTCHVSTLPETVPSVPTNIAFS
DVQSTSAFTLWIRPDTILGYFQNYKITTYLQRAQCKEWESEECVEYQKIQYLYEAHLTEETVYGLKFERWYR
FQVAASTNAGYGNASNWISTKTLPGPPDGPPENVHVVATSPFSISISWSEPAVITGPTCYLIDVKSVDNDEF
NISFIKSNENKTEIKDLEIFTRYSVVITAFGTGNISAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPD
EVTKFQLTSLPPSQPNNGNIQVYQALVYREDDPTAVQIHNLSIIQKTNTFVIAMLEGLKGGHTYNSIVYAVNS
AGAGPKVPMRITMDIKAPAPKTKPTPIYDATGKLLVTSTTITIRMPICYSDDHGPIKNVQVLVTETGAQH
DGNVTKWYDAYFNKARPYFTNEGFPNPPCTEGKTKFSGNEEIIYIIGADNACMIPGNEDKICNGPLPKPKQYL
FKFRATNIMGQFTSDSDYSDPVKTLGEGLSERTLEIILSVTLCILSIIILGTAIFAFARIRQKQEGGTYSPQ
DAEIIDTLKLDQLITVADLELKDERLRLLSYRSIKPISKKSFLQHVEELCTNNNLKFQEEFSELPKFLQ
DLSSTDADLPWNRARNRFPNIKPYNNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLPGTVGDFW
RMVWETRAKTLVMLTQCCEKGRIRCHQYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGCMTVR
QCNTAWPEHGVPENSAPLIHFVKLVRRASRAHDTTPIVHCSAGVGRGTGVFIALDHLTQHINDHDFVDIYGL
VAELRSERMCMVQNLAQYIFLHQCLDLLSNKGSNQPICFVNYSALQKMDSLDAMEGDVELEWEETTM

```

A search of sequence databases reveals that the NOV2c amino acid sequence has 1988 of 2301 amino acid residues (86%) identical to, and 2151 of 2301 amino acid residues (93%) similar to, the 2302 amino acid residue ptnr:SPTREMBL-ACC:O88488 protein from *Rattus norvegicus* (Rat) (Glomerular Mesangial Cell Receptor Protein-Tyrosine Phosphatase Precursor (EC 3.1.3.48)) (E = 0.0). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV2c is expressed in at least Synovium/Synovial membrane, Kidney. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CuraGen Acc. No. CG50718-05. The sequence is predicted to be expressed in the *Rattus norvegicus* :glomerular mesangial. because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AF063249|acc:AF063249.1) a closely related *Rattus norvegicus* glomerular mesangial cell receptor protein-tyrosine phosphatase precursor (PTPRQ) mRNA, complete cds homolog.

Homologies among each of the above NOV2 proteins will be shared by the other NOV2 proteins insofar as they are homologous to each other as shown below in Table 2G.

Any reference to NOV2 is assumed to refer to all three of the NOV2 proteins in general, unless otherwise noted.

Table 2G Alignment of NOV2a, b, and c

5		10	20	30	40	50	60								
	NOV2a	MDFLIIFLLFFIGTSETQVDVSNVVPGRYDITISSISTTYTSPVTRI	GASN--EPGPPV	58											
	NOV2b	MDFLIIFLLFFIGTSETQVDVSNVVPGRYDITISSISTTYTSPVTRI	VTINVT	1											
	NOV2c	MDFLIIFLLFFIGTSETQVDVSNVVPGRYDITISSISTTYTSPVTRI	VTINVT	60											
10		70	80	90	100	110	120								
	NOV2a	FLAGERVGSAGILLSWNTPPNPNNGRIISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTN	118												
	NOV2b	FLAGERVGSAGILLSWNTPPNPNNGRIISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTN	1												
15	NOV2c	FLAGERVGSAGILLSWNTPPNPNNGRIISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTN	120												
		130	140	150	160	170	180								
20	NOV2a	LNPGETTYEIKVAAENSAGIGVFSDPFLFQTAES	APGKVVDF	TCEAVP	FSS	KLMWYTS-	175								
	NOV2b	LNPGETTYEIKVAAENSAGIGVFSDPFLFQTAES	APGKVVDF	TCEAVP	FSS	KLMWYTS-	1								
	NOV2c	LNPGETTYEIKVAAENSAGIGVFSDPFLFQTAES	APGKVVDF	TCEAVP	FSS	KLMWYTS-	180								
		190	200	210	220	230	240								
25	NOV2a	ATKK	KITSFKISVKH	NRSGIVVKEVSIRVE	CILSAS	LPLHCN	ENSESFLWSTASPSPTLG	235							
	NOV2b	ATKK	KITSFKISVKH	NRSGIVVKEVSIRVE	CILSAS	LPLHCN	ENSESFLWSTASPSPTLG	1							
	NOV2c	QPNG	KITSFKISVKH	NRSGIVVKEVSIRVE	CILSAS	LPLHCN	ENSESFLWSTASPSPTLG	240							
		250	260	270	280	290	300								
30	NOV2a	RVTPPSRTTHSSSTLTQNEISSV	KEPISFVVTHLRPYTTYLFEVSAATEAGYIDSTIV	294											
	NOV2b	RVTPPSRTTHSSSTLTQNEISSV	KEPISFVVTHLRPYTTYLFEVSAATEAGYIDSTIV	1											
	NOV2c	RVTPPSRTTHSSSTLTQNEISSV	KEPISFVVTHLRPYTTYLFEVSAATEAGYIDSTIV	300											
35		310	320	330	340	350	360								
	NOV2a	RTPE	SVEGPPQNCV	TGNITGKSFS	SILWDPPTIVTGKFSYRVELYGPS	AGRILDN	STKDI	354							
	NOV2b	RTPE	SVEGPPQNCV	TGNITGKSFS	SILWDPPTIVTGKFSYRVELYGPS	AGRILDN	STKDI	1							
	NOV2c	RTPE	SVEGPPQNCV	TGNITGKSFS	SILWDPPTIVTGKFSYRVELYGPS	AGRILDN	STKDI	359							
40		370	380	390	400	410	420								
	NOV2a	KFAFTNLTPFTMYDVYIAAETSAGTGPKSNISVFTPPDVGAVFDLQLAEVESTQVRITW	414												
	NOV2b	KFAFTNLTPFTMYDVYIAAETSAGTGPKSNISVFTPPDVGAVFDLQLAEVESTQVRITW	1												
45	NOV2c	KFAFTNLTPFTMYDVYIAAETSAGTGPKSNISVFTPPDVGAVFDLQLAEVESTQVRITW	419												
		430	440	450	460	470	480								
50	NOV2a	KKPRQPNGIINQYRVKVLVPETGIILENTLLTGNNEINDEMAPEIVNIV	QPMVGLYEGSA	474											
	NOV2b	KKPRQPNGIINQYRVKVLVPETGIILENTLLTGNNEINDEMAPEIVNIV	QPMVGLYEGSA	1											
	NOV2c	KKPRQPNGIINQYRVKVLVPETGIILENTLLTGNNEINDEMAPEIVNIV	QPMVGLYEGSA	479											
		490	500	510	520	530	540								
55	NOV2a	EMSSDLHSLATFIYNSHPDKNFPARNRAEDQTS	PVVTT	RNQYITDIAAEQLTYV	LT	RLRR	534								
	NOV2b	EMSSDLHSLATFIYNSHPDKNFPARNRAEDQTS	PVVTT	RNQYITDIAAEQLTYV	LT	RLRR	1								
	NOV2c	EMSSDLHSLATFIYNSHPDKNFPARNRAEDQTS	PVVTT	RNQYITDIAAEQLTYV	LT	RLRR	539								
		550	560	570	580	590	600								
60	NOV2a	EWAE	TMGFS	RYTIM	SSASRDNI	TS	SPG	ELSA	QNF	RVTH	VIT	TEV	FLH	WDPPD	588
	NOV2b	EWAE	TMGFS	RYTIM	SSASRDNI	TS	SPG	ELSA	QNF	RVTH	VIT	TEV	FLH	WDPPD	1
	NOV2c	TEHM	ISVSA	ETIM	GEGPPTVLS	VRTR	QQV	ESSIK	IINYKN	NISS	SSIL	LYND	PPD	599	
65		610	620	630	640	650	660								

32

		1330	1340	1350	1360	1370	1380	
5	NOV2a						
	NOV2b	AFATKVGNGNQFSNVVKFTTQESVDPDVQNMQCMATSWQSVLVKWDPPKKANGIITQYMT						1366
	NOV2c	AFATKVGNGNQFSNVVKFTTQESVDPDVQNMQCMATSWQSVLVKWDPPKKANGIITQYMT						1377
		1390	1400	1410	1420	1430	1440	
10	NOV2a						
	NOV2b	VERNSTKVSPQDHMYTFIKLLANTSYVFKVRASTASAGEDESTCHVSTLPETVPSVPTNI						1426
	NOV2c	VERNSTKVSPQDHMYTFIKLLANTSYVFKVRASTASAGEDESTCHVSTLPETVPSVPTNI						1437
		1450	1460	1470	1480	1490	1500	
15	NOV2a						
	NOV2b	AFSDVQSTSATLTWIRPDTILGYFQNYKITTQLRAQCKEWESEECVEYOKIQYLYEAHL						1486
	NOV2c	AFSDVQSTSATLTWIRPDTILGYFQNYKITTQLRAQCKEWESEECVEYOKIQYLYEAHL						1497
		1510	1520	1530	1540	1550	1560	
20	NOV2a						
	NOV2b	TEETVYGLKKFRWYRFQVAASTNAGYGNASNWISTKTLPGPPDGPPEVHVHVATSPFISIS						1546
	NOV2c	TEETVYGLKKFRWYRFQVAASTNAGYGNASNWISTKTLPGPPDGPPEVHVHVATSPFISIS						1557
25		1570	1580	1590	1600	1610	1620	
	NOV2a						
	NOV2b	ISWSEPAVITGPTCYLIDVKSVDNDEFNISFIKSNEENKTIEIKDLEIFTRYSVVITAF						1606
	NOV2c	ISWSEPAVITGPTCYLIDVKSVDNDEFNISFIKSNEENKTIEIKDLEIFTRYSVVITAF						1617
30		1630	1640	1650	1660	1670	1680	
	NOV2a						
	NOV2b	GNISAAYVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKFQLTSLPPSQPNGNIQV						1666
	NOV2c	GNISAAYVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKFQLTSLPPSQPNGNIQV						1677
35		1690	1700	1710	1720	1730	1740	
	NOV2a						
	NOV2b	YQALVYREDDPTAVQIHNLSIIQKTNTFVIAMLEGLKGGHTYNISVYAVNSAGAGPKVPM						1726
	NOV2c	YQALVYREDDPTAVQIHNLSIIQKTNTFVIAMLEGLKGGHTYNISVYAVNSAGAGPKVPM						1737
40		1750	1760	1770	1780	1790	1800	
	NOV2a						
	NOV2b	RITMDIKAPARPKTKPTPIYDATGKLLVTSTTITIRMPICYSDDHGPIKNVQVLVTETG						1786
	NOV2c	RITMDIKAPARPKTKPTPIYDATGKLLVTSTTITIRMPICYSDDHGPIKNVQVLVTETG						1797
45		1810	1820	1830	1840	1850	1860	
	NOV2a						
	NOV2b	AQHDGNTWKYDAYFNKARPYFTNEGFPNPPCTEGKTKFSGNEEIIYGADNACMIPGNE						1846
	NOV2c	AQHDGNTWKYDAYFNKARPYFTNEGFPNPPCTEGKTKFSGNEEIIYGADNACMIPGNE						1857
50		1870	1880	1890	1900	1910	1920	
	NOV2a						
	NOV2b	DKICNGPLKPKKQYLFKFRATNIMGQFTSDSDYSDPVKTLGEGLSERTLEIILSVTLCLLS						1906
	NOV2c	DKICNGPLKPKKQYLFKFRATNIMGQFTSDSDYSDPVKTLGEGLSERTLEIILSVTLCLLS						1917
55		1930	1940	1950	1960	1970	1980	
	NOV2a						
	NOV2b	IILLGTAIFAFARIRQKQKEGGTYSQDAEIIIDTKLKLQDLITVADLELKDERLTR						1962
	NOV2c	IILLGTAIFAFARIRQKQKEGGTYSQDAEIIIDTKLKLQDLITVADLELKDERLTR						1977
60		1990	2000	2010	2020	2030	2040	
	NOV2a						
	NOV2b	-----FISKKSFQHVVEELCTNNLKFQEEFSELPKFLQDLSSTDADLPWNRKNRFPN						2017

35 The disclosed NOV2a polypeptide has homology to the amino acid sequences shown
in the BLASTP data listed in Table 2H.

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 12621078 ref NP_075214.1 (NM_022925)	protein tyrosine phosphatase, receptor type, Q [Rattus norvegicus]	2302	1893/2306 (82%)	2077/2306 (89%)	0.0
gi 125977 sp P16621 LAR_DROME	PROTEIN-TYROSINE PHOSPHATASE DLAR PRECURSOR (PROTEIN-TYROSINE-PHOSPHATE PHOSPHOHYDROLASE)	2029	410/1587 (25%)	680/1587 (42%)	1e-94
gi 10728878 gb AAF53837.2 (AE003663)	Lar gene product [Drosophila melanogaster]	2037	410/1587 (25%)	680/1587 (42%)	2e-94
gi 7290546 gb AAF45998.1 (AE003432)	Ptp4E gene product [Drosophila melanogaster]	1767	417/1645 (25%)	694/1645 (41%)	8e-94

gi 1362625 pir A49502	protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type 4E, splice form A precursor - fruit fly (Drosophila melanogaster)	1767	416/1645 (25%)	693/1645 (41%)	1e-92
-----------------------	---	------	----------------	----------------	-------

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 2I. In the ClustalW alignment of the NOV2 proteins, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 2I. ClustalW Analysis of NOV2

- 1) Novel NOV2a (SEQ ID NO:6)
- 2) gi|12621078|ref|NP_075214.1| (NM_022925) protein tyrosine phosphatase, receptor type, Q [Rattus norvegicus] (SEQ ID NO:37)
- 3) gi|125977|sp|P16621|LAR_DROME PROTEIN-TYROSINE PHOSPHATASE DLAR PRECURSOR (PROTEIN-TYROSINE-PHOSPHATE PHOSPHOHYDROLASE) (SEQ ID NO:38)
- 4) gi|10728878|gb|AAF53837.2| (AE003663) Lar gene product [Drosophila melanogaster] (SEQ ID NO:39)
- 5) gi|7290546|gb|AAF45998.1| (AE003432) Ptp4E gene product [Drosophila melanogaster] (SEQ ID NO:40)
- 6) gi|1362625|pir|A49502 protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type 4E, splice form A precursor - fruit fly (Drosophila melanogaster) (SEQ ID NO:41)

		10	20	30	40	50
25	NOV2A	MDFLIIFLLFIGTSETQVDVSNVVPCTRYDITISSIS--TTTSPVTR				
	gi 12621078	MDDFHFSFLFLIGTSESOVDVSSSFDGTGYDITLSSVSA--TTTSSPVSR				
	gi 125977	-----MGLQMTAARPIAALSLLVLSLLTWHTPTLVDAHPPEITRK				
	gi 10728878	-----MGLQMTAARPIAALSLLVLSLLTWHTPTLVDAHPPEITRK				
	gi 7290546	MDCATRKQQQLRAHHQQQIQIOTHTGRKROQLQKORHNHHHYTONSOQQ				
30	gi 1362625	MDCATRKQQQLRAHHQQQIQIOTHTGRKROQLQKORHNHHHYTONPOQQ				
		60	70	80	90	100
35	NOV2A	IGAS--NEPGPPVFLAGERVGSAGILLSWNTPPNPNGRIISYIVKYKEVC				
	gi 12621078	TLATNVTKPGPPVFLAGERVGSAGILLSWNTPPNPNGRIISYIVKYKEVC				
	gi 125977	PONQGVVRGGVASFYCAARGDPPPSIVWRKNGKKVSG-----				
	gi 10728878	PONQGVVRGGVASFYCAARGDPPPSIVWRKNGKKVSG-----				
	gi 7290546	QKHFVWLVLVGLITLFLAQHANADLVI--NVENASSNANAFYRTDYSPPF				
40	gi 1362625	QKHFVWLVLVGLITLFLARHANADLVI--NVENASSNANAFYRTDYSPPF				
		110	120	130	140	150
45	NOV2A	PWMOTVYTOVRSKPDSLEVLLTNINPGITYEIKVAAENNSAGIGVFSDPFL				
	gi 12621078	PWMOTAYTRARAKPDSLEVLLTNINPGITYEIKVAAENNSAGIGVFSDPFL				
	gi 125977	-T-OSRYTVLEQPGGISILRTEFVRACRDDAPYECVAENGVGDAVSADAT				
	gi 10728878	-T-OSRYTVLEQPGGISILRTEFVRACRDDAPYECVAENGVGDAVSADAT				
	gi 7290546	GFPEPNTTIPASDIG-KDIKFSRALPGTEYNFWLYYTNSTHREQLTWTVN				
	gi 1362625	GFPEPNTTIPASDIG-KDIKFSRALPGTEYNFWLYYTNSTHREQLTWTVN				

		160	170	180	190	200
	NOV2A				
	gi 12621078	FQTABSAFGKVVDFTEAVPFSSK-IMWYTS-ATKKKITSFKISVKHNRS				
5	gi 125977	FQTABSAFGKVVNLVEALNYSVNLIMWYLEROPNGKITSFKISVKHARS				
	gi 10728878	LTIIYEGWQ---KTATSGDKTPAGFFVITOGPGTRVIEVGHVLMTCKAIG				
	gi 7290546	ITTAIDPDP---ANLSVQLRSSKSAFILTWRPE---GSGRYSGFRIIRVLGLTD				
	gi 1362625	ITTAIDPDP---ANLSVQLRSSKSAFILTWRPE---GSGRYSGFRIIRVLGLTD				
10		210	220	230	240	250
	NOV2A				
	gi 12621078	GIVVKEVSIRVECIISASLFLHCNENSESFLWSTASPSPTLGRVTPPST				
	gi 125977	GIVVKDVSLSRVEDIISGKLE-ECNENSESFLWSTTSPSPTLGRVTPPTVRT				
15	gi 10728878	NETPNIIYWIKNQTKVDMNSNPRYSIKDGFLOIENSREEDOGKYECVAENSM				
	gi 7290546	NETPNIIYWIKNQTKVDMNSNPRYSIKDGFLOIENSREEDOGKYECVAENSM				
	gi 1362625	LPFERSYSLEGNETIQLSAK---ELTPGGSYQVQAYSVYOGKESVAYTSFN				
		260	270	280	290	300
20	NOV2A				
	gi 12621078	THSSSTLTQNEISSV-KEPISFVVTHLRPYTTYLFVSAATTEAGYIDST				
	gi 125977	TOSSSTAARSKISSVWKEPISFVVTHLRPYTTYLFVSAVTTEAGYIDST				
	gi 10728878	GTEHSAKATNLVVKVRVPPPTFSRPPETISEVMLGSNLSNCIAVGSMPH				
25	gi 7290546	GTEHSAKATNLVVKVRVPPPTFSRPPETISEVMLGSNLSNCIAVGSMPH				
	gi 1362625	FTTKPNTPGKIIVWFRNETTLLVLWQPPFPAGTYTHYRVSITPDDATOSV				
		310	320	330	340	350
30	NOV2A				
	gi 12621078	IVRTPESVVEGEPONCVTGNTEKSFSLIWDPPPTIVTGKFSYRVELYGPS				
	gi 125977	IVRTPESVVEGEPONCIMGNTGKAFSISWDPPPTIVTGKFSYRVELYGP-				
	gi 10728878	VKWMKGSDELTPENEMPIGRNVQLINIQESAN-----				
35	gi 7290546	VKWMKGSDELTPENEMPIGRNVQLINIQESAN-----				
	gi 1362625	LYVEREGEPPGPAIAAFKGLVPEREYNISVQT-----				
		360	370	380	390	400
40	NOV2A				
	gi 12621078	AGRILDNSTKDLKFAFTNLTPTMYDVYIAAETSAGTGPKSNISVFTTDP				
	gi 125977	SGRILDNSTKDLRFATHTLTPTMYDVYIAAETSAGVGPKSNISVFTTDP				
	gi 10728878	-----YTCIAASTLGQIDSVSVVKVQS				
45	gi 7290546	-----YTCIAASTLGQIDSVSVVKVQS				
	gi 1362625	-----VSEDETS-SVETTARYLTVER				
		410	420	430	440	450
50	NOV2A				
	gi 12621078	VPGAVFDLQIAEVESIQVRITWKKPRQPNGIINOYRVKVLVPEITGIIEN				
	gi 125977	VPGAVFDLQIAEVEATEIRITWKKPRQNGIISOYRVKVSILEITGVVLEN				
	gi 10728878	LEPTAPTDVQISEVTATSVRLWSYKG-----PEDLQYYVIQYKPKNAN				
	gi 7290546	LEPTAPTDVQISEVTATSVRLWSYKG-----PEDLQYYVIQYKPKNAN				
	gi 1362625	VLNVTFDEAYTTSSS--FRVRMEPPR---TYSEFDAYQVMLSTSRRTFN				
55		460	470	480	490	500
	NOV2A				
	gi 12621078	TLLTGNNI-INDPMAPFIVNIVQPMVGLYEGSAEMSSDLHSLATFIYNH				
	gi 125977	TLLTGDEISINPMSPFIMNLVDEMIGEYEGSGEMSSDLHSPASFIYNH				
60	gi 10728878	QAFSEISG---IITMYVVRALSPYTEYEFYVIAVNNIGRG-----				
	gi 7290546	QAFSEISG---IITMYVVRALSPYTEYEFYVIAVNNIGRG-----				
	gi 1362625	VPRAANGD---SVYFDYDILEPGRTYEVVVKTIADNVN-----				
		510	520	530	540	550
65	NOV2A				
	gi 12621078	PDKNFPAARNRAEDQTSFVVTTRNOYTIDIAAEQLTYVLIIRLRFAETMG				
	gi 125977	PHNDFPASTRAEEQSSFVVTTRNOYMTDITAEQLSYVVRRLVPFTEHTIS				
	gi 10728878	-----PFSAPATCTTGETKMESAP-----				
70	gi 7290546	-----PFSAPATCTTGETEMESAP-----				
		---SWPASGEVTLRPRVRSLG-----				

```

gi|1362625| ---SWPASGEVTLRPRVRLG-----
                    560      570      580      590      600
5  NOV2A      ....|.....|.....|.....|.....|.....|
   gi|12621078| FSRYTIMSSASRDNLTSFG---PLSAONFRVTHVTITIEVFLHDPDFPV
   gi|125977|   VSAFTIMGEGPPTVLTVRTREQVPSIIQIINYNKISSSSILLVWDPPPY
   gi|10728878|   -----RNVQVETLSSSTMVITWEPPETP
   gi|7290546|   -----RNVQVETLSSSTMVITWEPPETP
10 gi|1362625|   -----GFLDDR---SNALHISWEPAETG
   -----GFLDDR---SNALHISWEPAETG
                    610      620      630      640      650
15 NOV2A      FFHYYLITLIDVENQSKS---IILRTLNLSLVLIGLKKYTKYKMRVAA
   gi|12621078| NGKITHYTIYATELDTNR---AFQMTTVDNSFLITGLKKYTRKMRVAA
   gi|125977|   NGQVTGYKVYYTINSNQPEASWNSOMVDNSELTVSDVTPHAIYTVRVOA
   gi|10728878| NGQVTGYKVYYTINSNQPEASWNSOMVDNSELTVSELTPHAIYTVRVOA
   gi|7290546|   RQDSYRISYHECTNASEV---PAPFPVAAESQITTNLTETLDSILLAGR
20 gi|1362625|   RQDSYRISYHECTNASEV---PAPFPVAAESQITTNLTETLDSILLAGR
                    660      670      680      690      700
25 NOV2A      STHVGESSLSEENDIFVRTSEDEPESSPDVEVIDVTADEIRLKWSPPEK
   gi|12621078| STHVGESSLSEENDIFVRTPEDEPESSPDVQVGTGVSPELRLKWSPPEK
   gi|125977|   YTSMGAGPMSTP-----
   gi|10728878|   YTSMGAGPMSTP-----
   gi|7290546|   RYLLAVQALSKG-----
30 gi|1362625|   RYLLAVQALSKG-----
                    710      720      730      740      750
35 NOV2A      PNGIIAYEVLYKNIDTLYMKNTSTTDIILRNLRPHLYNISVRSYTRFG
   gi|12621078| PNGIIAYEVLYQNADTLFVKNTSTTDIISDLKKEYTLYNISIRSYTRLG
   gi|125977|   -----VQVKAQQGV-----
   gi|10728878|   -----VQVKAQQGV-----
   gi|7290546|   -----VASNASDIT-----R-YTRP-
   gi|1362625|   -----VASNASDIT-----R-YTRP-
                    760      770      780      790      800
40 NOV2A      HGNQVSSLLSVRTSESVDPDSAPENITYKNISSCEIETSELPPSSPENGIIQ
   gi|12621078| HGNQSSLLSVRTSETVPDSAPENITYKNISSCEIETSELPPSPENGIIQ
   gi|125977|   -----SQPSNFRATDICEATVTLQWTKPHSSSENIV
   gi|10728878|   -----SQPSNFRATDICEATVTLQWTKPHSSSENIV
45 gi|7290546|   -----AAPLIOELRSIDQE---LMLSWRSDVNSROD
   gi|1362625|   -----AAPLIOELRSIDQE---LMLSWRSDVNSROD
                    810      820      830      840      850
50 NOV2A      KHTIILKRSNG-NEERTINTTSLTONIKGLKKYTOYIIEVSASTLKGGGV
   gi|12621078| KYTIIYLKRSNS-HEARTINTTSLTQITGGLKKYTHYVIEVSASTLKGGGT
   gi|125977|   HYELIYNDTIYANQAAHKRISNSEAYTLDGLYPDTLYYIWLARSORGECA
   gi|10728878|   HYELIYNDTIYANQAAHKRISNSEAYTLDGLYPDTLYYIWLARSORGECA
   gi|7290546|   RYEVHYQRNGT-REERTMATNETSLTTHYLPFGSGYEVKVHAIHSH---GV
55 gi|1362625|   RYEVHYQRNGT-REERTMATNETSLTTHYLPFGSGYEVKVHAIHSH---GV
                    860      870      880      890      900
60 NOV2A      RSAPISILTEEDAPDSPPQDFSVKOLSGVTVKLSWQPF--LEPNCIILYY
   gi|12621078| RSRPISILTEEDAPDSPPQDFSVKOLSGVTVMLSWQPF--LEPNCIILYY
   gi|125977|   TTPPIPVRTKQYVPGAPERNITAIATSSTTISLSWLPPPVERSNGRIIYY
   gi|10728878|   TTPPIPVRTKQYVPGAPERNITAIATSSTTISLSWLPPPVERSNGRIIYY
   gi|7290546|   RSEPHSYFOAVEP--KPPONLTLOTVHTNLVVLHQAQ--EGSD--FSEYV
65 gi|1362625|   RSEPHSYFOAVEP--KPPONLTLOTVHTNLVVLHQAQ--EGSD--FSEYV
                    910      920      930      940      950
70 NOV2A      TVYVWR---SSLKTN-VTETSLSLSDLDYNVEYSAYVTASTREFGDGKT
   gi|12621078| TVYVWDK---SSLRAIN-ATASLVLSDLDYNVDYGACVTASTREFGDGNA
   gi|125977|   KVEFVYVGREDDEATTMTINMTSIVLDLKKRWTEYKTVVLGASVGDG-E

```

gi|10728878| KVEFVEVGREDDEATTMTLNMTSIVLDELKRWTEYKIVWLACTSVGDC-F
gi|7290546| VRYRTDA---SPWQRTISGLHENEARTKTMHGYERYLVQVNTVS-FGVESP
gi|1362625| VRYRTDA---SPWQRTISGLHENEARTKTMHGYERYLVQVNTVS-FGVESP

5 960 970 980 990 1000
NOV2A
gi|12621078| RSNITISFOTPEG-PSDEPKDVYYANLSSSSITLFWTPPS--KPNGIIQYY
gi|125977| RSSIINFRTPEGEPSPDNDVHYVNLSSSSITLFWTPPV--KPNGIIQYY
10 RSHPIILRTQEDVPGD-PQDVKATPLNSTSIHVSWKPLEKDRNGIIRGY
gi|10728878| RSHPIILRTQEDVPGD-PQDVKATPLNSTSIHVSWKPLEKDRNGIIRGY
gi|7290546| HPLELNVMTPEPQ----PVSNNVPLVDSRNLTLEWP----REDGHVDFY
gi|1362625| HPLELNVMTPEPQ----PVSNNVPLVDSRNLTLEWP----REDGHVDFY

15 1010 1020 1030 1040 1050
NOV2A
gi|12621078| SVYYRNTSGTFMQNFTLHEVTNDFDNMTVSTIIDKLTFISYYTFWLTAST
gi|125977| SVYYQNTSGTFVQNFLLQVTKESDNVTVSARIYRLAIFSYTFWLTAST
20
gi|10728878|
gi|7290546|
gi|1362625|

25 1060 1070 1080 1090 1100
NOV2A
gi|12621078| SVGNGNKSSDIEVYTDQDVPEGFVGNLTYESISSTAINVSWEPPAQPNG
gi|125977| SVGNGNKSSDIIHVYTDQDIPEGPVGNLTYESISSTAIHVSWEPPSQPNG
gi|10728878|HIHAQELRDEG
gi|7290546|HIHAQELRDEG
30 T-----LKKWETDEEDR
gi|1362625|T-----LKKWETDEEDR

35 1110 1120 1130 1140 1150
NOV2A
gi|12621078| LVFYVSLILOQTTP-RHVRPPLVTYERSIYFDNLEKYTDYILKITPSTER
gi|125977| LVFYVSLINLOQSPPRHMIPPLVTYENSIDFDLEKYTDYIFKITPSTER
gi|10728878| KGFLNEPFKEFVVD-----TLEFNVTGLQPDTKYSIOVAALTRK
gi|7290546| KGFLNEPFKEFVVD-----TLEFNVTGLQPDTKYSIOVAALTRK
40 VEFKNVTQLEDLSS-----P---SVRIPIEDLSPGRCYRFVQASSN-
gi|1362625| VEFKNVTQLEDLSS-----P---SVRIPIEDLSPGRCYRFVQASSN-

45 1160 1170 1180 1190 1200
NOV2A
gi|12621078| GFSDYYTAQLYIKTEEDIPETSPINTEFNLSSTSVLLSWDPPVKPNGAI
gi|125977| GFSEYTYTQLHIKTEEDVPDTPPIINTEFNLSSTSVLLSWDPPVKPNGAI
gi|10728878| CDGDRSAAIVKTPGCVFVRPTVSLKIMEREPIVSLLEWERPAQTYGEL
gi|7290546| CDGDRSAAIVKTPGCVFVRPTVSLKIMEREPIVSLLEWERPAQTYGEL
50 G-----IRSG-----T-----
gi|1362625| G-----IRSG-----T-----

55 1210 1220 1230 1240 1250
NOV2A
gi|12621078| ISYDITL---CGPNENYSFITSNDYILLEEISPFITLYSFFAAARTKGL
gi|125977| LGYHITL---CGPHANHTFVTSGNHTVLEEISPFITLYSFFAAARTKGL
gi|10728878| RGYRIRWGVKDCALKEEMLSGPQMTKKRFDNLERGVEYEFVAGSNHIGI
60 RGYRIRWGVKDCALKEEMLSGPQMTKKRFDNLERGVEYEFVAGSNHIGI
gi|7290546|THLSTRTMPLIQSDVFIANAGHEOQ
gi|1362625|THLSTRTMPLIQSDVFIANAGHEOQ

65 1260 1270 1280 1290 1300
NOV2A
gi|12621078| GPSSILFFYTDSEVLAAPPQNLTLINCTSDYVWLKWSSEPLPGGIVKVYS
gi|125977| GPSSILFFYTDSEVLAAPPQNLTLINCTSDYVWLKWSSEPLPGGIVKVYS
gi|10728878| GOETVKIFOTPEGTEGGPPSNITIRFOTPDVLCVTWDEPTRHRNGIITR
70 GOETVKIFOTPEGTEGGPPSNITIRFOTPDVLCVTWDEPTRHRNGIITR
gi|7290546|DETITLSYTFETPADSTRFDIYR
gi|1362625|DETITLSYTFETPADSTRFDIYR

75 1310 1320 1330 1340 1350
NOV2A
gi|12621078| FKIHEHETDTIYYKNISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNGNQF

gi|12621078| ~~SKLHE~~ETDTVFYKNISGLQTDAKLEGLEPVSTYSVSVSAFTKVGNGNQY
gi|125977| ~~YDVQFA~~-----
gi|10728878| ~~YDVQFA~~-----
gi|7290546| ~~ESMGD~~-----
gi|1362625| ~~ESMGD~~-----

5

10 NOV2A
gi|12621078| SNVVKFTTQESVPDVQNMQCMATSWQSVLVKWDPPKKANGIITQYMTV
gi|125977| SNVVEFTTQESVPEAVRNIECVARDWQSVSVRWDPKRKTNGIITHYMITV
gi|10728878| -----KK-----ID
gi|7290546| -----KK-----ID
gi|1362625| -----P-----TT
-----P-----TT

15

20 NOV2A
gi|12621078| ERNSTKVSPODHMYTFIKLLANTSYVFKVRASTSAEGDESTCHYSTLPE
gi|125977| GGNSTKVSPTOPTYTFIKLLPNTSYVFEVRASTSAEGNESRCDYSTLPE
gi|10728878| HGLGSEKNTLRLKAVETNLEENTETIFRVRAYTKQAGPFSDKLIVETER
gi|7290546| HGLGSEKNTLRLKAVETNLEENTETIFRVRAYTKQAGPFSDKLIVETER
gi|1362625| KDKEKLANDTERKLSFSGITPGKLYNVTWTVS-----GGVASLPVQRIYR
KDKEKLANDTERKLSFSGITPGKLYNVTWTVS-----GGVASLPVQRIYR

25

30 NOV2A
gi|12621078| TVPSVETNPAFSDVOSTSATLTWIRPDTILGYFQNYKITTLQRAQCKKEW
gi|125977| TVPSAPTINVAFSNVOSTSATLTWTKEDTIFGYFQNYKITTLQRAQCKREW
gi|10728878| DMGRAPMSLOEATSECTAEIWWPEVTSRGKLLGYKIFYMTAVE-----
gi|7290546| DMGRAPMSLOEATSECTAEIWWPEVTSRGKLLGYKIFYMTAVE-----
gi|1362625| LHPLPISDKAIOVAAREITLHWTAPAGEYTDDELOYLSADEEAP-----
LHPLPISDKAIOVAAREITLHWTAPAGEYTDDELOYLSADEEAP-----

35

40 NOV2A
gi|12621078| ESEECVEYOKILOYLYEAHLTEETVYGLKKERWYRFOVAASTNAGYGNASN
gi|125977| EPEECIEHOKDOYLYEANOETETVHGLKKERWYRFOVAASTNVGYNSASE
gi|10728878| -----DLDDWQTKTVGLTESADLVNLEKFAQYAVATAARFKNGLGRLSE
gi|7290546| -----DLDDWQTKTVGLTESADLVNLEKFAQYAVATAARFKNGLGRLSE
gi|1362625| -----OLLQN-VTKN--TEITLQGLRPYHNYTFTVVVRSGS--IQG--
-----OLLQN-VTKN--TEITLQGLRPYHNYTFTVVVRSGS--IQG--

45

50 NOV2A
gi|12621078| WISTKTLPGPPDGPPENVHVVATSPFISISWSEPAVITGPTCYLLIDVKS
gi|125977| WISTQTLPGPPDGPPENVHVVATSPFGINISWSEPAVITGPTRYLLIDVKS
gi|10728878| KVTVRK---PEDVPLNLRADVSTHSMTLWSPPPIRLT-PVNYKISFDA
gi|7290546| KVTVRK---PEDVPLNLRADVSTHSMTLWSPPPIRLT-PVNYKISFDA
gi|1362625| -----TDFADVSVSTLMRSSAPISASQTLTAPFGKVDYFQPSD-
-----TDFADVSVSTLMRSSAPISASQTLTAPFGKVDYFQPSD-

55

60 NOV2A
gi|12621078| VDNDEFNISFIKSNEENKTIEIKDLEIFTRYSVVITAFVGNVSRAYTDGK
gi|125977| VDDDDFNISFLKSNEENKTTEINNLEVFTRYSVVITAFVGNVSRAYTDGK
gi|10728878| MK-----
gi|7290546| MK-----
gi|1362625| VC-----
VC-----

65

70 NOV2A
gi|12621078| SSAEMIVTTLTLESAPKDEENNMTFOKIPDEVTKFQLTFLPESQPNENIQVY
gi|125977| SSAEVIITTLTLESVPKDEENNMTFOKIPDEVTKFQLTFLPESQPNENIRVY
gi|10728878| -----VFVDSQGSQT-----QIVPKREIILKH
gi|7290546| -----VFVDSQGSQT-----QIVPKREIILKH
gi|1362625| -----EGEVTREWS-----LEPAEQHCPTDYF
-----EGEVTREWS-----LEPAEQHCPTDYF

75

40

		2110	2120	2130	2140	2150
5	NOV2A	NEFIATQGPLPGTVGDFWRMVWETRAKTLVMLTCCFEKGRIRCHQYWPE				
	gi 12621078	NEFIATQGPLPGTVGDFWRMVWETRAKTLVMLTCCFEKGRIRCHQYWPE				
	gi 125977	NAYVATQGPLQETFDVDFWRMCWELKTATIVMMTRLEERTRIKCDQYWPTR				
	gi 10728878	NAYVATQGPLQETFDVDFWRMCWELKTATIVMMTRLEERTRIKCDQYWPTR				
	gi 7290546	REFIVTQGPLHSTREEFWRMCSNSRAIVMLTRCFEKGREKCDQYWPVD				
10	gi 1362625	REFIVTQGPLHSTREEFWRMCSNSRAIVMLTRCFEKGREKCDQYWPVD				
		2160	2170	2180	2190	2200
	NOV2A	NKPVTVEGDIVITKLMEDVQIDWTIRDLKIERH--GDCMTVRQCNETAWP				
15	gi 12621078	NKPVTVEGDIVITKLMEDVQIDWTIRDLKIERH--GDCMTVRQCNETAWP				
	gi 125977	G--TETYGQIFVTITETQELATYSIRTEQLCRQGFNDREIKOLOFTAWP				
	gi 10728878	G--TETYGQIFVTITETQELATYSIRTEQLCRQGFNDREIKOLOFTAWP				
	gi 7290546	R-VAMFYGDIKVQLITDTHYHDWSISEFMVSRN--CESRIMRHFHETTPW				
	gi 1362625	R-VAMFYGDIKVQLITDTHYHDWSISEFMVSRN--CESRIMRHFHETTPW				
20		2210	2220	2230	2240	2250
	NOV2A	DHGVPENSAPLTFEVKLVRASTRAHDTTETMIVHCSAGVGRGTGVFIALDHLT				
	gi 12621078	DHGVPENTTPLTFEVKLVRASTRAHDTTETMIVHCSAGVGRGTGVFIALDHLT				
25	gi 125977	DHGVPDHPAPFLOELRRCAITPPESGPIVHCSAGVGRGTGVFIALDHLT				
	gi 10728878	DHGVPDHPAPFLOELRRCAITPPESGPIVHCSAGVGRGTGVFIALDHLT				
	gi 7290546	DHGVPPEPQSLVRFVRAFRDVICTMRPIIVHCSAGVGRSGTFIALDRIL				
	gi 1362625	DHGVPPEPQSLVRFVRAFRDVICTMRPIIVHCSAGVGRSGTFIALDRIL				
30		2260	2270	2280	2290	2300
	NOV2A	QHINNDHDEVDIYGLVAELRSERMCMVQNLAQYIFLHQCILDLS-----				
	gi 12621078	QHINNDHDEVDIYGLVAELRSERMCMVQNLAQYIFLHQCILDLS-----				
	gi 125977	BRMKHEKIIDYGHVTCLEAQRNVMVOTEDQYIFIHDAILEALICG----				
35	gi 10728878	BRMKHEKIIDYGHVTCLEAQRNVMVOTEDQYIFIHDAILEALICG----				
	gi 7290546	QHITHKSQYVDIEGIVFAMRKERVEMVOTEQOYVCIHQCLLAVLEGKEHLL				
	gi 1362625	QHITHKSQYVDIEGIVFAMRKERVEMVOTEQOYVCIHQCLLAVLEGKEHLL				
40		2310	2320	2330	2340	2350
	NOV2A	-----NKGSNQPICFVNYSALQKMDSLDAMEGGDVELEWEETTM----				
	gi 12621078	-----NKGGHQPVCFVNYSALQKMDSLDAMEG-DVELEWEETTM----				
	gi 125977	--VTEVPARNLHTHLQKLLITEPGETISGMEVEFKKLSNVKMDSSKFVTA				
	gi 10728878	--VTEVPARNLHTHLQKLLITEPGETISGMEVEFKKLSNVKMDSSKFVTA				
45	gi 7290546	ADSLLELHANDGYEVTKIYLEROPOTKMGTLPTRASLAMEKLDADLMTNK				
	gi 1362625	ADSLLELHANDGYEVTKIYLEROPOTKMGTLPTRASLAMEKLDADLMTNK				
50		2360	2370	2380	2390	2400
	NOV2A	-----				
	gi 12621078	-----				
	gi 125977	NLPCKNKHKNRLVHILPYESSRVYLTPIHGIEGSDYVNASFIDGYRYSAY				
	gi 10728878	NLPCKNKHKNRLVHILPYESSRVYLTPIHGIEGSDYVNASFIDGYRYSAY				
	gi 7290546	DEDEDQEQQQQQQLQ-----LATEVKPKGSN				
55	gi 1362625	DEDEDQEQQQQQQLQ-----LATEVKPKGSN				
60		2410	2420	2430	2440	2450
	NOV2A	-----				
	gi 12621078	-----				
	gi 125977	IAAQGPVQDAABDFWRMLWEHNSTIVVMLTKLKEMGREKCFQYWPHERSV				
	gi 10728878	IAAQGPVQDAABDFWRMLWEHNSTIVVMLTKLKEMGREKCFQYWPHERSV				
	gi 7290546	DDEDEDEDDDDDDDDQPLNNETTATSSASCSSS-----THDVHV				
	gi 1362625	DDEDEDEDDDDDDDDQPLNNETTATSSASCSSS-----THDVHV				
65		2460	2470	2480	2490	2500
	NOV2A	-----				
	gi 12621078	-----				
70	gi 125977	RYQYYVVDPIAEYNMPQYKLRFKVTDARDGSSRTVRQFQFIDWPQGV				
	gi 10728878	RYQYYVVDPIAEYNMPQYKLRFKVTDARDGSSRTVRQFQFIDWPQGV				

gi 7290546	VLOEATEKPKQERICAGTQSHADTESDNTDSDDDDGDKGVAKDCAV
gi 1362625	VLOEATEKPKQERICAGTQSHADTESDNTDSDDDDGDKGVAKDCAV
	2510 2520 2530 2540 2550
5
NOV2A	-----
gi 12621078	-----
gi 125977	KSGEFIGIDFIGQVHKTKQFGQDGPITVHCSAGVGRSGVFITLSIVLERM
10	gi 10728878 KSGEFIGIDFIGQVHKTKQFGQDGPITVHCSAGVGRSGVFITLSIVLERM
gi 7290546	-----
gi 1362625	-----
	2560 2570 2580 2590
15
NOV2A	-----
gi 12621078	-----
gi 125977	QYEGVLDVFTVRILRSQRPAMVQTEDQYHFCYRAALEYLGSDNYTN
20	gi 10728878 QYEGVLDVFTVRILRSQRPAMVQTEDQYHFCYRAALEYLGSDNYTN
gi 7290546	-----
gi 1362625	-----

Tables 2J-2EE list the domain descriptions from DOMAIN analysis results against NOV2a. This indicates that the NOV2a sequence has properties similar to those of other proteins known to contain this domain.

Table 2J. Domain Analysis of NOV2a

gnl|Smart|smart00194, PTPc, Protein tyrosine phosphatase, catalytic domain (SEQ ID NO:93)
CD-Length = 264 residues, 99.6% aligned
Score = 318 bits (816), Expect = 2e-87

NOV 1:	1983	KFQEEFSELPK-FLQDLSSTADLPWNRAKNRFPNIKPYNNNRVKLIADASVPGSDYINA	2041
		+ + + + ++ ++	
30	Sbjct:	1 GLEEEFEKLQRLTPDDL SCTVAILPENRDKNRYKDVLPYDHTRVKL-KPPPGE GSDYINA	59
NOV 1:	2042	SYISGYLCPNEFIATQG PLPGTVGDFWRMVWETRAKTLVMLTQCFEKGRI RCHQYWPEDN	2101
		+ + + + +	
	60	SYIDGFNRPKAYIATQGPLSTVEDFWRMVWEEKVPVIVMLTELVEKGREKCAQYWPEKE	119
35	NOV 1:	2102 KPVTVFGDIVITKL MEDVQIDWTIRDLKIERHG--DCMTVRQC NFTA WPEHGV PENSAPL	2159
		+ + + + + + ++ + +	
	Sbjct:	120 GGS LTYGDITVTLKSVEKVDY TIR TLEV TNGGSE TRTVTHYHTN WPDHGV PESPKSL	179
40	NOV 1:	2160 IHFVKLV RASRAH--DTTPMIVHCSAGVGR TG VFIALDHLTQHINDHDFVDIYGLVAELR	2217
		+ + ++ ++ ++ + + + +	
	Sbjct:	180 LDLVRAVRKSQSTLRNSGPIVVHCSAGVGR TGTFIADILLQQL EAGKEVDIFEIVKELR	239
NOV 1:	2218	SERMC MVQNLAQYIFLHQILDLL	2241
		+ ++ +	
45	Sbjct:	240 SQRPGMVQTEEQYIFLYRAILEYL	263

Table 2K. Domain Analysis of NOV2a

gnl|Pfam|pfam00102, Y_phosphatase, Protein-tyrosine phosphatase (SEQ ID NO:94)

CD-Length = 235 residues, 100.0% aligned

Score = 275 bits (704), Expect = 2e-74

5	NOV 1:	2008	NRAKNRFPNIKPYNNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGFLPGTVGDF	2067
	Sbjct:	1	NKEKNRYKDVLPYDHTRVKL-KPLGDESDYINASYVDGYKKPKAYIATQGFLPNTIEDF	59
	NOV 1:	2068	WRMVWETRAKTLVMLTQCFEKGRIRCHQYWPEDNKPVTVEGDI-VITKLMEDVQIDWTIR	2126
	Sbjct:	60	WRMVWEEKVRVIVMLTELVEKGREKCAQYWPEKEGSLTYGDFTVTCVSVEKKKDDYTVR	119
10	NOV 1:	2127	DLKIERHGDC--MTVRQCNETAWPEHGVPEPENSAPLIHFVKLVASRAH-DTTPMIVHCSA	2183
	Sbjct:	120	TLELTNSGDDERTVVKHYHYTGWPDHGVPEPSPKSILDLRKVRKSKGTPDDGPVIVHCSA	179
15	NOV 1:	2184	GVGRTGVFIALDHLTQHINDHDFVDIYGLVAELRSEPMCMVQNLAQYIFLHQCILD	2239
	Sbjct:	180	GIGRTGTFFIAIDILLQOLEKEGVVDVFDTVKKLRSQRPQGMVQTEEQYIFYDAILE	235

Table 2L. Domain Analysis of NOV2a

gnl|Smart|smart00404, PTPc motif, Protein tyrosine phosphatase, catalytic domain motif (SEQ ID NO:95)

CD-Length = 105 residues, 100.0% aligned

Score = 120 bits (301), Expect = 8e-28

20	NOV 1:	2138	TVRQCNETAWPEHGVPEPENSAPLIHFVKLVASRAH--DTTPMIVHCSAGVGRGTGVFIALD	2195
	Sbjct:	1	TVKHYHYTGWPDHGVPEPSPDSILEFLRAVKKSLNKSANNGPVVVHCSAGVGRGTGTFVAID	60
25	NOV 1:	2196	HLTQHI-NDHDFVDIYGLVAELRSEPMCMVQNLAQYIFLHQCILD	2239
	Sbjct:	61	ILLQOLEAGTGEVDIFDIVKELRSQRPQAVQTEQYLYFLYRALLE	105

Table 2M. Domain Analysis of NOV2a

gnl|Pfam|pfam00041, fn3, Fibronectin type III domain (SEQ ID NO:96)

CD-Length = 86 residues, 100.0% aligned

Score = 60.8 bits (146), Expect = 8e-10

30	NOV 1:	54	PGPPVFLAGERVGSAGILLSWNTPPNPNNGRIISYIVKYKEVCPWMQTVYTQVRSKPDLSLE	113
	Sbjct:	1	PSAPTNLTVTDVTSTSLTSLWSPPPDGNGPITGYEVEYQPVNS--GEEWNEITVPGTTTS	58
35	NOV 1:	114	VLLTNLNPGETTYEIKVAAENSAGIGVFS	141
	Sbjct:	59	YTLTGLKPGTEYEVVRQAVNGGGNGPPS	86

Table 2N. Domain Analysis of NOV2a

gnl|Pfam|pfam00041, fn3, Fibronectin type III domain (SEQ ID NO:96)
 CD-Length = 86 residues, 95.3% aligned
 Score = 58.9 bits (141), Expect = 3e-09

NOV 1: 659 SSPQDVEIDVTADEIRLKWSPPEKPNGIIAYEVLYKNIDTLYMKNT-----STTDIIL 713
 |+| ++ | |||+ + | ||| || | ||| |+ +++ | +|| |
 Sbjct: 2 SAPTNLTVTDVTSTSLTLSWSPPPDGNGPITGYEVEYQPVNSGEEWNEITVPGTTTSYTL 61

NOV 1: 714 RNLRPHTLYNISVRSYTRFGHG 735
 |+| | | + |++ |+|
 Sbjct: 62 TGLKPGTEYEVVRVQAVNGGGNG 83

Table 2O. Domain Analysis of NOV2a

gnl|Pfam|pfam00041, fn3, Fibronectin type III domain (SEQ ID NO:96)
 CD-Length = 86 residues, 100.0% aligned
 Score = 57.0 bits (136), Expect = 1e-08

NOV 1: 1330 PDVVQNMQCMATSWQSVLVKWDPPKKANGIITQYMVTV-----ERNSTKVSPQDHMYT 1382
 | |+ + |+ + | || || || | | | | | |
 Sbjct: 1 PSAPTNLTVTDVTSTSLTLSWSPPPDGNGPITGYEVEYQPVNSGEEWNEITVPGTTTSYT 60

NOV 1: 1383 FIKLLANTSIVFKVRASTSAGEGDES 1408
 | | | +|+| | | |
 Sbjct: 61 LTGLKPGTEYEVVRVQAVNGGGNGPPS 86

Table 2P. Domain Analysis of NOV2a

gnl|Pfam|pfam00041, fn3, Fibronectin type III domain (SEQ ID NO:96)
 CD-Length = 86 residues, 98.8% aligned
 Score = 53.1 bits (126), Expect = 2e-07

NOV 1: 753 SAPENITYKNISSGEIELSFLPPSSPNGIIQKYTIYLKRSNGNE---ERTINTTSLTQNI 809
 ||| |+| +++| +||+ || || | | + + | | |+ |+ + +
 Sbjct: 2 SAPTNLTVTDVTSTSLTLSWSPPPDGNGPITGYEVEYQPVNSGEEWNEITVPGTTTSYTL 61

NOV 1: 810 KGLKKYTOYIIIEVSASTLKGEVRS 834
 ||| |+| + | | | | |
 Sbjct: 62 TGLKPGTEYEVVRVQAVNGGGNGPPS 86

Table 2Q. Domain Analysis of NOV2a

gnl|Pfam|pfam00041, fn3, Fibronectin type III domain (SEQ ID NO:96)
 CD-Length = 86 residues, 95.3% aligned
 Score = 52.4 bits (124), Expect = 3e-07

NOV 1: 848 SPPQDFSQKQLSGVTVKLSWQPPLEPNGIILYYTVYVWR-----SSLKTINV--TETSLEL 901
 | | + +| ++ ++ ||| || + || | | | | | | | |
 Sbjct: 2 SAPTNLTVTDVTSTSLTLSWSPPPDGNGPITGYEVEYQPVNSGEEWNEITVPGTTTSYTL 61

NOV 1: 902 SDLDYNVEYSAYVTASTRFGDG 923
 + | || | | |+|
 Sbjct: 62 TGLKPGTEYEVVRVQAVNGGGNG 83

Table 2R. Domain Analysis of NOV2a

gnl|Pfam|pfam00041, fn3, Fibronectin type III domain (SEQ ID NO:96)
CD-Length = 86 residues, 91.9% aligned
Score = 51.6 bits (122), Expect = 5e-07

```
NOV 1: 1148  TFKNLSSTSVLLSWDPVPKPNGAIISYDLTLOGPNENYSFIT-----SDNYIILEELSPF 1202
             |  +++|||+ ||| || | |++ | |  +  +  |  ||
Sbjct: 8      TVTDVSTSTSLTILSWSPFPDGNPITGYEVEYQPVNSGEEWNEITVPGTTTSYTLTGLKPG 67

NOV 1: 1203  TLYSFFAAARTRKGLGPSS 1221
             ||  |  || |||
Sbjct: 68    TEYEVVRVOAVNGGGNGPPS 86
```

Table 2S. Domain Analysis of NOV2a

gnl|Pfam|pfam00041, fn3, Fibronectin type III domain (SEQ ID NO:96)
CD-Length = 86 residues, 94.2% aligned
Score = 51.2 bits (121), Expect = 6e-07

NOV 1:	1235	PPQNLTILINCTSDFVWLKWSPSPPLGGIVKVYSFK-IHEHETDTIYYKNISGFKTEAKLV	1293
		+ + + + + + + +	
Sbjct:	3	APTNLTVTDVTSLSLTLSWSPFPDNGPITGYEVEYQPVNSGEEWNEITVPGTTTSYTLT	62
NOV 1:	1294	GLEPVSTYSIRVSFAFTKVNG	1314
		+ + +	
Sbjct:	63	GLKPGTEYEVVRVQAVNGGGNG	83

Table 2T. Domain Analysis of NOV2a

gnl|Pfam|pfam00041, fn3, Fibronectin type III domain (SEQ ID NO:96)
CD-Length = 86 residues, 100.0% aligned
Score = 49.7 bits (117), Expect = 2e-06

NOV 1:	1420	PSVPTNIAFSDVQSTSATLTWIRPDTILGYFONYKITTQLRAQCKKEWEESEECVEYQKIQ	1479
		+ + + + +	
Sbjct:	1	PSAPTNLTVTDVSTSLTSLWSPPDPGNGPITGYEVEYQ-----PVNSGEEWNEITV--	52
NOV 1:	1480	YLYEAHLTEETVYGLKKFRWYRFQVAASTNAGYGNAS	1516
		+ +	
Sbjct:	53	---PGTTTSYTLTGLKPGTEYEVRRVOAVNGGGNGPPS	86

Table 2U. Domain Analysis of NOV2a

gnl|Pfam|pfam00041, fn3, Fibronectin type III domain (SEQ ID NO:96)
CD-Length = 86 residues, 98.8% aligned
Score = 47.4 bits (111), Expect = 9e-06

NOV 1:	940	DPPKDVYYANLSSSSIIILFWTPPSKPNIGIIQYYSVYYRNT-SGTFMQNFTLHEVTNDFDN	998
		++ ++ + + ++ + +	
Sbjct:	2	SAPTNLTVTDVTSTSLTLSWSPPPDGNGPITGYEVEYQPVNSGEEWNEITVPGTTT----	57
NOV 1:	999	MTVSTIIDKLTIFFSYTFLWLTASTSVGNNGNKS	1030

Sbjct: 58 | + | + | + | ||| |
 ---SYTLTGLKPGTEYEVVRVQAVNGGGNGPPS 86

Table 2V. Domain Analysis of NOV2a

gnl|Pfam|pfam00041, fn3, Fibronectin type III domain (SEQ ID NO:96)
 CD-Length = 86 residues, 91.9% aligned
 Score = 47.0 bits (110), Expect = 1e-05

5 NOV 1: 1530 GPPENVHVVATSPFISISISWSEPAVITGP-TCYLIDVKSVNDDEFNISFIKSNEENKTIE 1588
 | |+| + |+++||| | || | ++ + |++ | +
 Sbjct: 2 SAPTNLTVTDVTSTSLTSLWSPPPDGNGPITGYEVEYQPVNSGEEWNEITVPGTTT-SYT 60

10 NOV 1: 1589 IKDLEIFTRYSVVITAFITGN 1608
 + |+ | | | + | |
 Sbjct: 61 LTGLKPGTEYEVVRVQAVNGG 80

Table 2W. Domain Analysis of NOV2a

gnl|Pfam|pfam00041, fn3, Fibronectin type III domain (SEQ ID NO:96)
 CD-Length = 86 residues, 96.5% aligned
 Score = 46.6 bits (109), Expect = 2e-05

15 NOV 1: 1633 DFPNNMTFQKIPDEVTKFQLTFLPPSQPNIGNIQVYQALVYREDDPTAVQIHNLSIIQKTN 1692
 | |+| + | |++ || || | |+ + +
 Sbjct: 2 SAPTNLTVTDVTS--TSLTSLWSPPPDGNGPITGYEVEYQPVNSGEEWNEITVPGTTTS- 58

20 NOV 1: 1693 TFVIAMLEGLKGGHTYNISVYAVNSAGAGP 1722
 | ||| | | + | ||| | ||
 Sbjct: 59 ----YTTLTGLKPGTEYEVVRVQAVNGGGNGP 84

Table 2X. Domain Analysis of NOV2a

gnl|Pfam|pfam00041, fn3, Fibronectin type III domain (SEQ ID NO:96)
 CD-Length = 86 residues, 98.8% aligned
 Score = 44.7 bits (104), Expect = 6e-05

25 NOV 1: 303 GPPQNCVTGNITGKSFSILWDPPTIVTGKFS-YRVELY---GPSAGRILDNSTKDLKFAF 358
 | | ++| | ++ | || | + || + +
 Sbjct: 2 SAPTNLTVTDVTSTSLTSLWSPPPDGNGPITGYEVEYQPVNSGEEWNEITVPGTTTSYTL 61

30 NOV 1: 359 TNLTPFTMYDVYIAAETSAGTGPKS 383
 | | | | |+| + | | || |
 Sbjct: 62 TGLKPGTEYEVVRVQAVNGGGNGPPS 86

Table 2Y. Domain Analysis of NOV2a

gnl|Pfam|pfam00041, fn3, Fibronectin type III domain (SEQ ID NO:96)
 CD-Length = 86 residues, 100.0% aligned
 Score = 43.1 bits (100), Expect = 2e-04

NOV 1: 561 PLSAQNFRVTHVTITEVFLHWDPPDPVF--FHHYLITILDVENQSKSIILRTLNSLSLVL 618
 | + | || || | + | || | + | + + + +
 Sbjct: 1 PSAPTNLTVTDVTSTSLTSLWSPPPDGNGPITGYEVEYQPVNSGEEWNEITVPGTTTSYT 60

NOV 1: 619 I-GLKKYTKYKMRVAASTHVGESSL 643
+ ||| |+|+|| | | |
Sbjct: 61 LTGLKPGTEYEVVRVQAVNGGGNGPPS 86

Table 2Z. Domain Analysis of NOV2a

gnl|Pfam|pfam00041, fn3, Fibronectin type III domain (SEQ ID NO:96)
CD-Length = 86 residues, 93.0% aligned
Score = 38.5 bits (88), Expect = 0.004

NOV 1: 1047 VGNLTYESISSTAINVSWVPPAQPNGLVFYY-VSLILQOTPRHVRPPLVT-YERSIYFDN 1104
||| ++|++ +|| || || + | | + | |
Sbjct: 4 PTNLTVTDTSTSLTSLWSPPDGNPITGYEVEYQPVNSGEEWNEITVPGTTTSYTLTG 63
NOV 1: 1105 LEKYTDYILKITPSTKGF 1124
|+ |+| +++ |
Sbjct: 64 LKPGTEYEVVRVQAVNGGGNG 83

Table 2AA. Domain Analysis of NOV2a

gnl|Smart|smart00060, FN3, Fibronectin type 3 domain; One of three types of internal repeat within the plasma protein, fibronectin. The tenth fibronectin type III repeat contains a RGD cell recognition sequence in a flexible loop between 2 strands. Type III modules are present in both extracellular and intracellular proteins. (SEQ ID NO:97)
CD-Length = 83 residues, 96.4% aligned
Score = 54.7 bits (130), Expect = 6e-08

NOV 1: 54 PGPPVFLAGERVGSAGILLSWNTPPNP-NGRIISYIVKYKEVCPWMQTVYTQVRSKPDSL 112
| || | | | + ||| ||+ | |+| |+|+| | + | +
Sbjct: 1 PPSPNLRLVTDVTSTSVTLSEWPPDDITGYIVGYRVEYREEGEWKEVNVTP----SSTT 56
NOV 1: 113 EVLLTNLNPGTTYEIKVAAENSAG 136
|| | ||| || +| | |
Sbjct: 57 SYTLTGLKPGTEYEFVRVANGA 80

Table 2BB. Domain Analysis of NOV2a

gnl|Smart|smart00060, FN3, Fibronectin type 3 domain; One of three types of internal repeat within the plasma protein, fibronectin. The tenth fibronectin type III repeat contains a RGD cell recognition sequence in a flexible loop between 2 strands. Type III modules are present in both extracellular and intracellular proteins. (SEQ ID NO:97)
CD-Length = 83 residues, 92.8% aligned
Score = 52.8 bits (125), Expect = 2e-07

NOV 1: 659 SSPQDVEVIDVTADEIRLKWSPPEKP-NGIIIAYEVLYKNID---TLYMKNTSTTDIILR 714
| | ++ | |||+ + | | | | |+| | |+ + + +||| |
Sbjct: 2 SPSPNLRLVTDVTSTSVTLSEWPPDDITGYIVGYRVEYREEGEWKEVNVTPSSTTSYTLT 61
NOV 1: 715 NLRPHTLYNISVRSYTR 731
|+| | | ||+
Sbjct: 62 GLKPGTEYEFVRVANG 78

Table 2CC. Domain Analysis of NOV2a

gnl|Smart|smart00060, FN3, Fibronectin type 3 domain; One of three types of internal repeat within the plasma protein, fibronectin. The tenth fibronectin type III repeat contains a RGD cell recognition sequence in a flexible loop between 2 strands. Type III modules are present in both extracellular and intracellular proteins. (SEQ ID NO:97)

CD-Length = 83 residues, 94.0% aligned
Score = 45.4 bits (106), Expect = 3e-05

NOV 1: 1235 PPQNLTLINCTSD FVWLKWSPLPGGIVKVYSFKIHEHETDTIYYKNISGFKTEAKLVG 1294
| | | + + | | | | | | | | + | + | | |
Sbjct: 3 PPSNLRVTDVTSTSVTLSEWPPDDITGIYGYRVEYREEGEWKVNVPSTTSYTTLTG 62

NOV 1: 1295 LEPVSTYSIRVSAFTKVG 1312
|+| + | |||
Sbjct: 63 LKPGTEYEFVRVAVNGEA 80

Table 2DD. Domain Analysis of NOV2a

gnl|Smart|smart00060, FN3, Fibronectin type 3 domain; One of three types of internal repeat within the plasma protein, fibronectin. The tenth fibronectin type III repeat contains a RGD cell recognition sequence in a flexible loop between 2 strands. Type III modules are present in both extracellular and intracellular proteins. (SEQ ID NO:97)

CD-Length = 83 residues, 100.0% aligned
Score = 42.7 bits (99), Expect = 2e-04

```
NOV 1: 561 PLSAQNFRVTHVTITEVFLHWDPPDPVFHHLITILDVENQSKSIILRTINS--LSLVL 618
          |   | ||| || | | |+||      + +  ++  + +  +   |   | |
Sbjct:  1  PSPPSNLRVTDVTSTSVTLSEWPPPPDDITGYIVGYRVEYREEGEWKEVNVTPSSSTSYTL 60
```

```
NOV 1: 619 IGLKKYTKYKMRVAASTHVGESS 641
          ||| |+|+ || |
Sbjct: 61  TGLKPGTEYEFVRVAVNGEAGEG 83
```

Table 2EE. Domain Analysis of NOV2a

gnl|Smart|smart00060, FN3, Fibronectin type 3 domain; One of three types of internal repeat within the plasma protein, fibronectin. The tenth fibronectin type III repeat contains a RGD cell recognition sequence in a flexible loop between 2 strands. Type III modules are present in both extracellular and intracellular proteins. (SEQ ID NO:97)

CD-Length = 83 residues, 92.8% aligned
Score = 41.2 bits (95), Expect = 7e-04

```
NOV 1: 848  SPPQDFSVKQLSGVTVKLSWQPPLEP-NGIILYYTVVWRSS----LKTINVTTLSLELS 902
          ||| + | ++ +| |||+| + | |+| | + + || |+
Sbjct: 2     SPPSNLRVTDVTSTSVTLSEPPDDITGYIVGYRVEYREEGEWKEVNVTPSSTTSYTLT 61
```

NOV 1: 903 DLDYNVEYSAYVTASTR 919
 | | | | |
Sbjct: 62 GLKPGTEYEFRVRAVNG 78

Receptor tyrosine phosphatases (rPTPs) are part of the signaling cascades that control cell survival, proliferation and differentiation. The novel protein tyrosine phosphatase described in the application contains a phosphatase domain and thirteen fibronectin type III repeats. It closely resembles rPTP-GMC1, a rat membrane phosphatase that is expressed in kidney glomerulus and is upregulated in response to kidney injury (Wright et.al. J Biol Chem 1998 Sep 11;273(37):23929-37). Tissue specificity of PTPs varies widely ; for eg rPTP-GMC1 is expressed by mesangial cells in the kidney while GLEPP1 (another membrane phosphatase) is expressed by podocytes in the kidney (Thomas et. al. ; J Biol Chem 1994 Aug 5;269(31):19953-62). Tappia et. al. demonstrated expression of a PTP in the liver could regulate the activity of the insulin and EGF receptors (Tappia et. al.; Biochem J 1993 May 15;292 (Pt 1):1-5). A number of phosphatases have been demonstrated to play a role in cancer, for eg. PTP zeta; a membrane phosphatase; is expressed in brain and is also expressed by a glioblastoma cell line (Krueger et. al.; Proc Natl Acad Sci U S A 1992 Aug 15;89(16):7417-21); rPTP alpha is expressed in breast tumors and correlates with tumor grade (Ardini et. al.; Oncogene 2000 Oct 12;19(43):4979-87). This phosphatase (rPTP alpha) is also expressed by human prostate cancer cell lines, oral squamous cell carcinoma and was correlated with histological grade of the oral tumor (Zelivianski et. al.; Mol Cell Biochem 2000 May;208(1-2):11-8; Berndt et al.; Histochem Cell Biol 1999 May;111(5):399-403). PTP-1B has been suggested to play a role in diabetes and obesity (Kennedy et. al.; Biochem Pharmacol 2000 Oct 1;60(7):877-83) while mutations in a PTP named EPM2A have been suggested as the cause of Lafora's disease (and autosomal recessive form of progressive myoclonus epilepsy) (Minassian et. al. Nat Genet 1998 Oct;20(2):171-4). Given the wide ranging effects of this family of proteins, we hypothesize that the novel protein described in this application plays a role in cancer, neurological, immune and metabolic diseases.

The disclosed NOV2 nucleic acid of the invention encoding a Protein tyrosine phosphatase precursor-like protein includes the nucleic acid whose sequence is provided in Table 2A, 2C, or 2E or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 2A, 2C, or 2E while still encoding a protein that maintains its Protein tyrosine phosphatase precursor like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications

include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

5 In the mutant or variant nucleic acids, and their complements, up to about 16 percent of the bases may be so changed.

The disclosed NOV2 protein of the invention includes the Protein tyrosine phosphatase precursor-like protein whose sequence is provided in Table 2B, 2D, or 2F. The invention also includes a mutant or variant protein any of whose residues may be changed from the
10 corresponding residue shown in Table 2B, 2D, or 2F while still encoding a protein that maintains its Protein tyrosine phosphatase precursor-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 18 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or
15 (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this Protein tyrosine phosphatase precursor-like protein (NOV2) may function as a member of a "Protein tyrosine phosphatase precursor family". Therefore, the NOV2 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various
20 pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to)
25 those defined here.

The NOV2 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in cancer including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the Protein tyrosine phosphatase precursor-like protein (NOV2) may be useful in gene therapy, and the Protein
30 tyrosine phosphatase precursor-like protein (NOV2) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cancer, kidney cancer, trauma, regeneration (*in vitro* and *in vivo*), viral/bacterial/parasitic infections, nephrological diseases including diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis,

glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, Lesch-Nyhan syndrome, Hirschsprung's disease, Crohn's Disease, appendicitis, or other pathologies or conditions. The NOV2 nucleic acid encoding the Protein tyrosine phosphatase precursor-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV2 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV2 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV2 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated NOV2 epitope is from about amino acids 1 to 100. In another embodiment, a NOV2 epitope is from about amino acids 200 to 300. In further embodiments, a NOV2 epitope is from about amino acids 450 to 500, from about amino acids 600 to 900, from about amino acids 950 to 1000, from about amino acids 1200 to 1300, from about amino acids 1400 to 1600, from about amino acids 1800 to 1900, from about amino acids 1950 to 2050, and from about amino acids 2200 to 2300. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV3

A disclosed NOV3 nucleic acid of 4538 nucleotides (also referred to as 134899552_EXT) encoding a novel human homolog of the *Drosophila* pecanex-like protein is shown in Table 3A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 101-103 and ending with a TGA codon at nucleotides 4439-4441. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 3A, and the start and stop codons are in bold letters.

Table 3A. NOV3 nucleotide sequence (SEQ ID NO:11).

<p><u>CATGAAGGAAAAATTCGAGTATTCTAATGGCTTTTAAATAATCATTATTGCTAGGTAAGTCTCTTC</u> <u>TACGCTGTATGAGACTGGTGGCTGTGATATGTCACTTGTGAATTTGAACCAGCAGCAAGAAGAGCATCCAA</u> <u>TATCTGGGACACAGATTCTCATGTATCCAGTTCTACCTCAGTTCGATTTATCCACATGATGTGATTGCGATT</u> <u>GAATAGACTATTGACCATTTGATACAGATTTGTTGGAGCAACAGGACATTGATCTAAGCCCTGACTTGGCAGC</u> <u>TACTTACGGCCCAACAGAAGAAGCTGCCAAAAGGTTAAACACTATTATCGCTTTTGGATCCTACCCAGCT</u> <u>GTGGATTGGCATTAACTTTGACAGACTCACACTTTTGGCCCTGTTTGATAGGAATCGTGAGATCCTGGAAAA</u> <u>TGTGTTAGCTGTCATCCTGGCTATTCTCGTGGCCTTTTGGGATCTATTCTTCTCATACAAGGATTCTTCAG</u> <u>AGATATCTGGGTCTTCCAGTTCTGCCTCGTCATAGCCAGCTGTCAATACTCACTGCTTAAGAGTGTTCAACC</u></p>
--

```

AGATTCTTCTTCTCCAGACATGGTCATAATCGTATCATTGCCTACAGTAGACCAGTTTATTTCTGCATATG
TTGCCGCTCTTATTTGGCTCTTGGATTATGGTAGCAGAAACCTGACTGCAACCAAGTTCAAATTATATGGAAT
AACTTTACCAATCCACTGGTGTTTATATCAGCCAGGGATTAGTTATAGTGTTCACACTCTGTTCCCAAT
AGTGTTTTTTCATTGGTCTCCTGCCTCAGGTGAATACATTTGTAATGTACCTTTGTGAACAATTGGATATTC
TATTTTTGGTGGTAATGCCACTACAAGCCTGCTTGACAGCACTTACAGTTTTATCTGTAGCATTGTTGCAGT
AGCCTTATTGTATGGATTATGTTATGGGGCTTACAGGATTCTTGGGATGGCCAGCATATCCAGTACTTTT
CTCCATTTTTTGTGGTTTATTAGTGGCAGTGTCTTACCATCTCAGCCGACAAAGCAGTGATCCATCTGTACT
TAGCTCTTTTAGTGCAATCCAAGATTTTCCAAAAACGGAAGAGAAAAATCCAGAAGACCCTCTATCTGAAGT
AAAAGATCCACTGCCTGAAAACTTAGAAATCTGTTAGTGAGCGATTACAGTCTGACCTGGTAGTATGCAT
TGTAATTGGTGTGCTGATTTTGCTATTATGTAAGCAGAGTCTTACAGTATTCAGCCTGCCCTCAAGTA
TGTGTTGTATACATTGGTTGGCTTTTGGGGTTTGTAAACCATTATGTGCTGCCTCAAGTTAGAAAACAGCT
ACCATGGCACTGTTTCTCATCTCTGCTAAAGACACTAGAGTATAATCAGTATGAAGTTCGAGATGCAGC
CACTATGATGTGGTTTGAGAACTTCATGTGTGGCTTCTTTTGTGGAGAAGAAATATAATCTATCCATTGAT
TGTTCTCAATGAAGTGAAGCAGTGCAGAGACAATTGCTAGTCCAAAGAACTGAATACAGAGTTAGGTGC
TTTAATGATCACTGTGTGCTGGTTTGAAGTTGCTACGATCCTCTTTTAGCAGCCCTACATATCAGTATGTTAC
AGTCATCTTACTGTGCTGTTTTTCAAATTTGACTATGAAGCTTTTTCAGAGACCATGCTGTTGGATCTCTT
CTTTATGTTCCACTTCCCAGCTTGATAGAAATCCAGGTTCCAGATGACAACAATCTGAATCCATCTTTATGA
CCCATGGCAGATCACATGGGGTTCTGCTTTCCATGCTTTTGTCTCAGCCTTTTGCAGTGCCTCGTTTCAGCCAT
GCTGTTTATTCAGGCTGCTGCTCTCGGCCCTTCTCTCTACTCCACTGAACCCCTTTCTGGGAAGTGCAATATT
CATCACTTCATATGTCGACCTGTGAATTCGAGGAGAGACTATAGCACAACAGAGTGGATCATTCAAA
TACCAAGATTGCTTCCCAGCTTGATAGAAATCCAGGTTCCAGATGACAACAATCTGAATCCATCTTTATGA
GCATTTAAGTATGCTTACAGCAGCCTCTGTGGTGATTGCTACTAGGACGGTGGGGAACTACAGTAC
AGGGGACTGTTTCATCCTTGCTCTGACTATCTCAATGCATTAGTACACCTTATAGAGATAGGCAATGGTCT
GGTCACTTTTCAGCTGCGGGGACTTGAATTCAGAGGTACCTACTGTCAACAACGGGAAGTGGAGGCCATTAC
TGAAGGTGTAGAGGAAGATGAAGATTTTGTGTTGTGAACCTGGCCATATTCTCACATGCTTTCATTTAA
TGCTGCATTTAGCCAGCGATGGCTAGCTTGGGAAGTGATAGTACACAAGTACATTCTGGAGGGTTATAGCAT
CACTGATAACAGTGTGCTTCTATGCTTCAAGTCTTTGATCTTCGGAAGTACTCACCCTTACTATGTCAA
GGGTATCATTATTATGTTACGACCTCGTCTAAGCTAGAGGAGTGGCTAGCTAATGAGACAATGCAGGAAGG
ACTTCGTCTGTGCTGATCGCAATTATGTCGATGTGGACCCGACCTTAAATCCAACATTGATGAAGACTA
TGACCACCGACTGGCAGGATATCTAGGGAGAGTTTCTGTGTGATTACCTCACTGGATAGAGTACTGCTC
TTCCCGAAGAGCAAAGCCTGTGGATGTGGACAAAGATTCAATCCCTAGTGAAGTCTCTGTTATGGACTCTGTGT
TCTGGGACGGAGAGCTTTGGGGACTGCATCCCATCATATGTCCAGTAAATTTAGAGTCAATTCCTCTATGGATT
GCATGCCCTATTTAAAGGAGATTTCCGTATTTCTTCAATTCGAGATGAATGGATCTTGTCTGACATGGAATT
GCTAAGAAAAGTAGTAGTCCCTGGGATCCGTATGTCCATTAACTTCATCAGGATCATTTTACTTCTCCAGA
TGAATATGATGACCTACTGTGCTCTATGAAGCCATAGTATCTCATGAGAAGAACCTCGTAATAGCCCATGA
AGGGGACCCCTGCATGGCGGAGTGCACTACTGCCAATCTCCTCCTGCTTGTCTGCTGCGGCATGTCTATGGA
TGATGGCACCATGAATATAAAATCATCATGCTCAACAGACGCTACCTGAGCTTCAGGCTCATTAAGTGAA
TAAGGAATGTGTCGAGTCTTTGGGCAGGGCAACAGCAGGAGCTTGTTTTCTACGTAACCGTAACCCAGA
GAGAGGTAGCATCCAAAATGCAAAGCAAGCCCTGAGAAACATGATAAATCATCTTGTGTATCAACCTATTGG
CTACCCAATCTTTGTCTCACCCTGACAACCTTCTTACTCTGACAGCCACGAACAGCTTAAAGACATTCTTGG
GGGTCTATCAGCTTGGGAAATATCAGGAACCTCATAGTGTCAACCTGGCACAGGCTTAGGAAAGGTTGCGG
AGCTGGATGTAAAGTGGTGGCAATATTGAAGATTCTGATACTGGAGGTGGGACTTCTTGCATGGTAACAA
TGCAACAACCTGCCAACATCCCCACAGCAACGTGACCCAGGGAAGCATTGGAATCCTGGGCAGGGATCAGG
AACTGGACTCCACCACCTGTACATCTTATCCTCCAACACTAGGTACTAGCCACAGCTCTCACTCTGTGCA
GTCCGGCCTGGTGCAGAGTCTCCTGCCCGGCTCAGTAGCCAGCCAGTCTTCTACTGCTATAGCAGCCG
GCATTCATCCCTCCGATGTCCACCCTGGGTTTGTGCTTGTGCGGCTCTTCTACTAGTGCATATCGCT
TCCAAACTTGCATCCATCCAAATCCCGACTGTGATGGTGAACCAATGGAACCTCAGGTGAGAGCGG
CCTGGCCTGTGTGCAGCAGGCTGCTTCTCCAGCAGCTCCAGCCAAAGCATCCAGCCTGCAAACATCA
CACTCTCGTGGCTTTCTTGCAGCAGAGGGAGGTGAGAGCAGTGCCACTGATGCACAGCCAGGCAACACCTT
AAGTCTGCCAACAAATTCACACTCCAGAAAGGCAGAGTGATTACAGAGTCCAAATGTGGATCCCAGTCA
AATTCTGGAAGGGATCAACCTGTCTAAAAGGAAGAGCTACAGTGGCCTGATGAAGGAATCCGGTTAAAGC
TGGGAGAAATAGCTGGAAGACTGGAGTCCGAGGAGGGCATGGAAGGCCATGTGATTACCGATGGGTGCC
TTGCAGCAGAGATCCAGTACCAGATCCACATCGACAAGGCAGTGCTTCTGGTCCAGATTGATGATAAATA
TGTGACTGTAATTGAACTGGGGTACTAGAATTTGGGGCTGAAGTGTGAGCCAGTGTATTATAAAGACAT
TTCTTTTTTCCCTCTCAATTCGAAGGCATTGGAAAAAGAGAGGAACAAGCAGAGATGCTGCAGGTATCACT
TT

```

The disclosed NOV3 nucleic acid sequence, localized to chromosome 14, has 2277 of 2283 bases (99%) identical to a gb:GENBANK-ID:AB018348|acc:AB018348.1 mRNA from *Homo sapiens* (*Homo sapiens* mRNA for KIAA0805 protein, partial cds) (E = 0.0).

5

A NOV3 polypeptide (SEQ ID NO:12) encoded by SEQ ID NO:11 has 1446 amino acid residues and is presented using the one-letter code in Table 3B. Signal P, Psort and/or Hydropathy results predict that NOV3 does not contain a signal peptide and is likely to be

localized to the plasma membrane with a certainty of 0.8000. In other embodiments, NOV3 may also be localized to the mitochondrial inner membrane with a certainty of 0.4714, the Golgi body with a certainty of 0.4000, or the endoplasmic reticulum (membrane) with a certainty of 0.3000.

5

Table 3B. Encoded NOV3 protein sequence (SEQ ID NO:12).

```
MSLVNFEPAARRASNIWDTDSHVSSSTSVRFYPHDVIRLNRLTIDTDLLEQQDIDLSPDLAATYGPTEEAA
QKVKHYYRFWILPQLWIGINFDRLLTLLALFDRNREILENVLAIVLAILVAFLGSILLIQGFFRDIWVFQFCL
VIASCQYSLKSVQPDSSSPRHGHNRIIAYSREPVYFCICCGLIWLLDYGSRLTATKFKLYGITFTNPLVFTI
SARDLVIVFTLCFFPIVFIFIGLLPQVNTFVMYLCEQLDIHIFGGNATTSLLAALYSFICSIVAVALLYGLCYG
ALQDSWDGQHIPVLFSTFCGLLVAVSYHLSRQSSDPSVLSLVQSKIIFPKTEEKNPEDPLSEVKDPLPEKLR
NSVSERLQSDLVVCIVIGVLYFAIHVSTVFTVLQPAKYVLYTLVGVGVFVTHYVLPQVRKQLPWHCFSHPL
LKTLEYNQYERDAATMMWFELHVVLLFVEKNIYPLIVLNELSSSAETIASPKKLNTLALMITVAGLK
LLRSSFSPTYQYVTVIFTVLFFKFDYEAFFSETMLLDLFFMSILFNKLWELLYKLQFVYTYIAPWQITWGSA
FHAFAPQFAVPRSAMLFIAAASAFFSTPLNPFGLSAIFITSYVRPVKFWERDYSTKRVDSNTRLASQLDR
NPGSDDNNSIFYEHLTRSLQHSKCGDLLLGRWGNYSTGDCFILASDYLNALVHLIEIGNGLVTFQLRGLG
FRGTYCQOREVEAITEGVEDEGFCCCEPFGHI PHMLSFNAAFQORLAWEVIVTKYILEGYSTIDNSAASML
QVFDLRKVLFTYYVIGIYVFTTSSKLEEWLANETMQEGLRLCADRNYVDVDPFTFNPNIDEDYDRLAGISR
ESFCVIYLNWIEYCRRRAKPDVDKSSLVTLCYGLCVLGRALGTASHHMSNLESFLYGLHALFKGDFR
ISSIRDEWIFADMELLRKVVPGRMSIKLHQDHFTSPDEYDDPTVLYEAIVSHEKNLVIAHEGDPAWRSVA
LANSPSLALRHVMDGTNEYKIIMLNRRYLSFRVIKVNKECVRGLWAGQQQLVFLNRNRPERSIQNAKQ
ALRNMINSSCDQPIGYPIFVSPLTTSYSDSHEQLKDILGGPISLGNIRNFIVSTWHLRLKGCAGCNSGGNI
EDSDTGGGTCTGNNATTANNPHSNVTQGSIGNPGQSGTGLHPPVTSYPPTLGTSHSSHVSQSLVRQSPA
RASVASQSSYCYSSRHSSLRMSTTGFPVPCRRSSTSQISLRNLPSSIQSRLSMVNQMEPSGQSGLACVQHGLP
SSSSSSQSIIPACKHHTLVGFLATEGGQSSATDAQPGNTLSPANNSHSRKAIEYIRVQIVDPSQILEGINLSK
RKELQWPDEGIRLKAGRNSWKDWSPQEGMEGHVIRHWVPCSRDPGTRSHIDKAVLLVQIDDKYVTVIETGVL
ELGAEV
```

The disclosed NOV3 amino acid sequence has 1355 of 1446 amino acid residues (93%) identical to, and 1409 of 1446 amino acid residues (97%) similar to, the 1446 amino acid residue ptnr:SP TREMBL-ACC:Q9QYC1 protein from *Mus musculus* (Mouse) (PECANEX 1) (E = 0.0).

NOV3 is expressed in at least Pancreas, Parathyroid Gland, Thyroid, Mammary gland/Breast, Ovary, Placenta, Uterus, Colon, Liver, Bone Marrow, Lymphoid tissue, Spleen, Tonsils, Prostate, Testis, Brain, Lung, and Kidney. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, NOV3 is predicted to be expressed in *Homo sapiens* heart, melanocyte, B-cells, larynx, skin, CNS, and multiple sclerosis lesions because of the expression pattern of the following sequences (which are publicly available ESTs for the sequence of the invention)

AB018348, BE881203, BE867469, BE867415, AB007895, NM_014801, U74315, BE880986, W500099, AW250617, AA426168, AW246742, AA284182, W46420, H14491, Z44921, BE930588, AI922381, AI215559, AA923742, AA582883, BE797814,

N75143, BE049421, F07632, BE797239, AI168579, AV653955, BE065657, AL079849, and BE767656, closely related *Homo sapiens* mRNA for KIAA proteins, partial cds homolog.

NOV3 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 3C.

5

Table 3C. BLAST results for NOV3					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
ref XP_027243.1 (XM_027243)	hypothetical protein XP_027243 [Homo sapiens]	619	619/619 (100%)	619/619 (100%)	0.0
gi 15076843 gb AAK8 2958.1 AF233450_1 (AF233450)	pecanex-like protein 1 [Homo sapiens]	2341	1372/1451 (94%)	1376/1451 (94%)	0.0
gi 6650377 gb AAF21 809.1 AF096286_1 (AF096286)	pecanex 1 [Mus musculus]	1446	1296/1446 (89%)	1344/1446 (92%)	0.0
gi 13171105 gb AAK1 3590.1 AF154413_1 (AF154413)	pecanex [Takifugu rubripes]	1703	1079/1466 (73%)	1204/1466 (81%)	0.0
gi 7290294 gb AAF45 755.1 (AE003423)	pcx gene product [alt 1] [Drosophila melanogaster]	3437	320/554 (57%)	424/554 (75%)	0.0

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 3D.

10

Table 3D. ClustalW Analysis of NOV3

15

20

25

30

- 1) NOV3 (SEQ ID NO:12)
- 2) ref|XP_027243.1| (XM_027243) hypothetical protein XP_027243 [Homo sapiens] (SEQ ID NO:42)
- 3) gi|15076843|gb|AAK82958.1|AF233450_1 (AF233450) pecanex-like protein 1 [Homo sapiens] (SEQ ID NO:43)
- 4) gi|6650377|gb|AAF21809.1|AF096286_1 (AF096286) pecanex 1 [Mus musculus] (SEQ ID NO:44)
- 5) gi|13171105|gb|AAK13590.1|AF154413_1 (AF154413) pecanex [Takifugu rubripes] (SEQ ID NO:45)
- 6) gi|7290294|gb|AAF45755.1| (AE003423) pcx gene product [alt 1] [Drosophila melanogaster] (SEQ ID NO:46)

	10	20	30	40	50	60
NOV3					
ref XP_027243.1	-----	-----	-----	-----	-----	1
gi 15076843 gb	MGSQTLQILRQGVWAALSGGWYD	PHQATFVNALHLYLWLFLGLPFTLYMALPSTMIIV	60			
gi 6650377 gb A	-----	-----	-----	-----	-----	1
gi 13171105 gb	MGSQTLQILRQGVWASVTGGWYD	PDQNTFVNALHLYIWLFLCFPFTLYMALQPSMVIV	60			
gi 7290294 gb A	-----	-----	-----	-----	-----	1
	70	80	90	100	110	120

55

56

	1030	1040	1050	1060	1070	1080		
NOV3	LAV	LAI	LVAF	LGS	LLI	QGFFR	DIWVFQFCLVIASCOYSL	172
ref XP_027243.1	LAV	LAI	LVAF	LGS	LLI	QGFFR	DIWVFQFCLVIASCOYSL	1
gi 15076843 gb	LAV	LAI	LVAF	LGS	LLI	QGFFR	DIWVFQFCLVIASCOYSL	106
gi 6650377 gb A	LAV	V	LAI	LVAF	LGS	LLI	QGFFR	172
gi 13171105 gb	LAV	V	LAV	LVAF	LGS	LLI	HGFFA	476
gi 7290294 gb A	LAV	V	LAV	LVAF	LGS	LLI	HGFFA	1

1090 1100 1110 1120 1130 1140
 NOV3 AYSRPVYFCI CCGLIWLLD YGSRNLTATKFKLYGTFETNP LVEISARDLVIVFTLCFPIV 232
 ref|XP_027243.1 1
 gi|15076843|gb| AYSRPVYFCI CCGLIWLLD YGSRNLTATKFKLYGTFETNP LVEISARDLVIVFTLCFPIV 112
 gi|6650377|gb|A AYSRPVYFCI CCGLIWLLD YGSRNLTTSKFKLYGVTFETNP LVLSSARDLVIVFTLCFPIV 232
 gi|13171105|gb| AYSRPVYFCI CCGLIWLLH YGSLRHTTSRFTLYGVALTSS LVLASARDLVIVFTLCFPII 536
 gi|7290294|gb|A 1

[illegible]

	1210	1220	1230	1240	1250	1260	
NOV3	SWDGQHLPVLFSI	FCGLLVAVSYHLSRQSSDPSVLSSI	VOSKIFPKTEEKNPEDPLSEVK				352
ref XP_027243.1							1
gi 15076843 gb	SWDGQHLPVLFSI	FCGLLVAVSYHLSRQSSDPSVLFSI	VOSKIFPKTEEKNPEDPLSEVK				124
gi 6650377 gb A	SWDGQHLPVLFSV	FCGLLVAVSYHLSRQSNDPSVLFSIMOSKI	FPKADEKNPEDPLSEVK				352
gi 13171105 gb	NTEPHHLPVLFSV	FCGLLVAVSYHLSRQSDPSVLMYVPLSKVF	FLRSKNPEDPLSEVK				656
gi 7290294 gb A							1

	1270	1280	1290	1300	1310	1320	
NOV3	DPLPEKLRLNSVSRERLQSDLVVCIVIGVLYFAIHVSTVFTVLPQALRYVLYLTIVGFVGFVT						412
ref XP_027243.1	DPLPEKLRLNSVSRERLQSDLVVCIVIGVLYFAIHVSTVFTVLPQALRYVLYLTIVGFVGFVT						1
gi 15076843 gb	DPLPEKLRLNSVSRERLQSDLVVCIVIGVLYFAIHVSTVFTVLPQALRYVLYLTIVGFVGFVT						130
gi 6650377 gb A	DPLPEKLRLNSVSRERLQSDLVVCIVIGVLYFAIHVSTVFTVLPQALRYVLYLTIVGFVGFVT						412
gi 13171105 gb	DPLPEKLRLNSVSRERLQSDLVVCIVIGVLYFAIHVSTVFTVLPQALRYVLYLTIVGFVGFVT						716
gi 7290294 gb A	DPLPEKLRLNSVSRERLQSDLVVCIVIGVLYFAIHVSTVFTVLPQALRYVLYLTIVGFVGFVT						1

	1330	1340	1350	1360	1370	1380	
NOV3						
ref XP_027243.1	HYVLPQVRKQLPWEFCFSHPLLKTLLEYNQYEVNRDAATMMWFEKLHVWLLFVEKNITLYPLIV						472
gi 15076843 gb	HYVLPQVRKQLPWEFCFSHPLLKTLLEYNQYEVNRDAATMMWFEKLHVWLLFVEKNITLYPLIV						1
gi 6650377 gb A	HYVLPQVRKQLPWEFCFSRPLLRRTADHSQYEVNRDAATMMWFEKLHVWLLFVEKNITLYPLIV						136
gi 13171105 gb	HYVLPQVRKQLPWEFCFSRPLLRRTADHSQYEVNRDAATMMWFEKLHVWLLFVEKNITLYPLIV						472
gi 7290294 gb A	HYVLPQLRKQLPWYCFSSHPLLKTKLEYQYEVNRDAAHVMWFEKLFHVWLLFVEKNITLYPLIV						776
						1

	1390	1400	1410	1420	1430	1440	
NOV3	LNELSS	SAETTIASPKKLNTEL	GALMITVAGLKLRRSSF	SSPTYQYVTVI	FTVLFFKF	DYE	532
ref XP_027243.1							1
gi 15076843 gb	LNELSS	SAETTIASPKKLNTEL	GALMITVAGLKLRRSSF	SSPTYQYVTVI	FTVLFFKF	DYE	142
gi 6650377 gb A	LNELSS	SAETTIASPKKLDTEL	GALMITIAGLKLRRSSF	SSSPTYQYITVI	FTVLFFKF	DYE	532
gi 13171105 gb	LNELSS	SAELASPKRLDTEVG	ALMITVAGLKLRRSSYS	SPTYQYITIL	FTVLFFTL	DHR	836
gi 7290294 gb A							1

1450 1460 1470 1480 1490 1500
 NOV3 AFSETMLLDLFFMSILFNKLWELLYKLOFVYTYTAPWQITWGSAFHAFAPFAVPRSA⁵⁹²
 ref|XP_027243.1 -----
 1

gi|15076843|gb| AFSETMLLDLFFMSILFNKILWELLYKLFQVYTYTAPWQITWGSFAHFAQPPFAVPHSAML 1487
gi|6650377|gb|A AFSETMLLDLFFMSILFSSKLWELLYKRFQVYTYVAPWQITWGSFAHFAQPPFAVPHSAML 592
gi|13171105|gb| HLSETLLDLFLMSIIFSKMWELLYKLFHVFYTYTAPWQITWGSFAHFAQPPFAVPHSAML 896
gi|7290294|gb|A ----- 1

5
110
15
10
15
20
25
30
35
40
45
50
55
60
65
70

NOV3
ref|XP_027243.1
gi|15076843|gb|
gi|6650377|gb|A
gi|13171105|gb|
gi|7290294|gb|A

1510 1520 1530 1540 1550 1560

FTQAVSAFSTPLNPFLLGSAIFITSYVRPVKFWERDYSTKRVDSNTRLASQLDRNPGS 652
----- 1
FTQAVSAFSTPLNPFLLGSAIFITSYVRPVKFWERDYNTKRVDSNTRLASQLDRNPGS 1547
FVQAVSAFSTPLNPFLLGSAIFITSYVRPVKFWERDYNTKRVDSNTRLASQLDRNPGS 652
FVQAVSSIFSTPLNPFLLGSAIFITSYVRPVKFWERDYNTKRVDSNTRLASQLDRNPGS 956
----- 47
MSTEESPDSEYTSAPVPVDCRVTDLK-----ENEMKQVDFDEDTRVLLVKN-----

1570 1580 1590 1600 1610 1620

DDNNLNSIFYEHLTRSLQHSLLCGDLLGRWGNYSTGDCFILASDYLNALVHLIEIGNGLV 712
----- 1
DDNNLNSIFYEHLTRSLQHSLLCGDLLGRWGNYSTGDCFILASDYLNALVHLIEIGNGLV 1607
DDNNLNSIFYEHLTRSLQHSLLCGDLLGRWGNYSTGDCFILASDYLNALVHLIEIGNGLV 712
DDNNLNSIFYEHLTRSLQHSLLCGDLLGRWGNYSTGDCFILASDYLNALVHLIEIGNGLV 1016
DRLAVGAKCTHYGAPLOT---GALGLGRVCPWHGACPNLENGDLEDFP-----GLD 97

1630 1640 1650 1660 1670 1680

TFQLRGLEFRGTYCQOREVEAITEGVEEDEGFCCEPFGHPHMLSFNAAFSQRWLAVEVT 772
----- 1
TFQLRGLEFRGTYCQOREVEAITEGVEEDEGFCCEPFGHPHMLSFNAAFSQRWLAVEVT 1667
TFQLRGLEFRGTYCQOREVEAITEGVEEDEGFCCEPFGHVPVLSFNAAFQRWLAVEVV 772
TFQLRGLEFRGTYCQOREVEAITEGVEEDEGFCCEPFGHPHILSFNAAFQRWLAVEVV 1076
SLP-----CYRVE 105

1690 1700 1710 1720 1730 1740

VTKYILEGYSITDNSAASMLQVFDLRKVLTTYVVKGIYYVTTSSKLEEWLANETMOEGL 832
----- 5
VTKYILEGYSITDNSAASMLQVFDLRKVLTTYVVKGIYYVTTSSKLEEWLANETMOEGL 1727
VTKYILEGYSITDNSAASMLQVFDLRRLVTTYVVKGIYYVTTSSKLEEWLANETMOEGL 832
VTKYVLEGYSITDNSAASMLQVFDLRRLTTYVVKGIYYVTTSSKLEEWLANETMOEGL 1136
VG-----NEGQ-----VMLRAKRSDDLNNKRLKNMV 131

1750 1760 1770 1780 1790 1800

RLCADRNYVDVDPTEFNPN-IDEDYDHRLAGISRESFCVIYLNWIEYCSSRRRAKPVVDVKD 891
RLCADRNYVDVDPTEFNPN-IDEDYDHRLAGISRESFCVIYLNWIEYCSSRRRAKPVVDVKD 64
RLCADRNYVDVDPTEFNPN-IDEDYDHRLAGISRESFCVIYLNWIEYCSSRRRAKPVVDVKD 1786
RLCADRNYVDVDPTEFNPN-IDEDYDHRLAGISRESFCVIYLSWIEYCSSRRRAKPLVDVKD 891
RGCSESNYVDLDATFNPN-IDEDYDHRISGISRDSFCGVYLGWIOYCNRRRTKPLDSEKD 1195
RRKPDQRFVIVVGGGPGSAGAVAVETIROEGTGRLIIVFCREDYLPYDRVKISKAMNLEIE 191

1810 1820 1830 1840 1850 1860

SSLVTLGYGLCVLGRRALGTASHMSSNLESFLYGLHALFKGDFRISSIRDEWIFADMEL 951
SSLVTLGYGLCVLGRRALGTASHMSSNLESFLYGLHALFKGDFRISSIRDEWIFADMEL 124
SSLVTLGYGLCVLGRRALGTASHMSSNLESFLYGLHALFKGDFRISSIRDEWIFADMEL 1846
SSLVTLGYGLCVLGRRALGTASHMSSNLESFLYGLHALFKGDFRISSVRDEWIFADMEL 951
SALVLLCFGLCVLGRRALGTAAHQMSSNLESFLYGLHALFKGDFRISSVRDEWIFADMEL 1255
Q-----LRFR-----DEEYKEYDTLWQGVAAEKLDTAQKELHCSNGYVVKYDKI 237

1870 1880 1890 1900 1910 1920

LRKVVVPGIRMSIK-LHQDHFTSPDEYDDPTVLYEAIVSHEKNLVIAHEGDPAWRSAYLA 1010
LRKVVVPGIRMSIK-LHQDHFTSPDEYDDPTVLYEAIVSHEKNLVIAHEGDPAWRSAYLA 183
LRKVVVPGIRMSIK-LHQDHFTSPDEYDDPTVLYEAIVSHEKNLVIAHEGDPAWRSAYLA 1905
LRKVVVPGIRMSIK-LHQDHFTSPDEYDDPTVLYEAIVSHEKNLVIAHEGDPAWRSAYLA 1010
LRKVVVPGIRMSIK-LHQDHFTSPDEYDDPTVLYEAIVSHEKNLVIAHEGDPAWRSAYLA 1314
YLATGCSAFRPFIFGVNLENVRVRELAOTKALLASITPESR-----VVC 282

1930 1940 1950 1960 1970 1980

NOV3		NSPSLLALRHVMDGDTNEYKIIMLNRRYLSFRVIKVNKECVRLWAGQQQELVFLNRNP	1070
ref XP_027243.1		NSPSLLALRHVMDGDTNEYKIIMLNRRYLSFRVIKVNKECVRLWAGQQQELVFLNRNP	243
gi 15076843 gb		NSPSLLALRHVMDGDTNEYKIIMLNRRYLSFRVIKVNKECVRLWAGQQQELVFLNRNP	1965
gi 6650377 gb A		NSPSLLALRHVMDGDTNEYKIIMLNRRYLSFRVIKVNKECVRLWAGQQQELVFLNRNP	1070
5	gi 13171105 gb	NSPSLLALRHVMDGDTNEYKIIMLNRRYLSFRVIKVNKECVRLWAGQQQELVFLNRNP	1374
	gi 7290294 gb A	LGSSEIALAAAGLVSKVSVTVVGRENVPLKAAFGAEIGQFVLQLEFEDNKVVMRMESG-	341
		1990 2000 2010 2020 2030 2040	
10 NOV3		ERGSIQNAKQALRNMINSSCDQPIGYPIFVSPLTTSYSDSHDOLKDI LGGPISLGNIRNF	1130
ref XP_027243.1		ERGSIQNAKQALRNMINSSCDQPIGYPIFVSPLTTSYSDSHDOLKDI LGGPISLGNIRNF	303
gi 15076843 gb		ERGSIQNAKQALRNMINSSCDQPIGYPIFVSPLTTSYSDSHDOLKDI LGGPISLGNIRNF	2025
gi 6650377 gb A		ERGSIQNAKQALRNMINSSCDQPIGYPIFVSPLTTSYSDSHDOLKDI LGGPISLGNIRNF	1130
gi 13171105 gb		ERGSIQNAKQALRNMINSSCDQPIGYPIFVSPLTTSYSDSHDOLKDI LGGPISLGNIRNF	1434
15	gi 7290294 gb A	-IAEIVGNEDGKVSFVVLVDITRIPCDLLILGTGSKLNTQFLAKSGVKVNRNGSVDPVDF	400
		2050 2060 2070 2080 2090 2100	
20 NOV3		IVSTWHRLRKGCCAGCNSGGNIEDSDTGGGTSCGTGNATTANNPHSNVTGGSIGNPGQGS	1190
ref XP_027243.1		IVSTWHRLRKGCCAGCNSGGNIEDSDTGGGTSCGTGNATTANNPHSNVTGGSIGNPGQGS	363
gi 15076843 gb		IVSTWHRLRKGCCAGCNSGGNIEDSDTGGGTSCGTGNATTANNPHSNVTGGSIGNPGQGS	2085
gi 6650377 gb A		IVSTWHRLRKGCCAGCNSGGNIEDSDTGGGTSCGTGNATTANNPHSNVTGGSIGNPGQGS	1190
gi 13171105 gb		VVSTWHRLRKGCCAGCNSGGNIEDSDAGC	1463
25	gi 7290294 gb A	LES--N-VP-----DVIYVGGDIANAHIHG	421
		2110 2120 2130 2140 2150 2160	
30 NOV3		CTGLHPPVTSYPPTLTGSHSSHSVQSGLVROSPARASVASQSS-YCYSS-RHSSLRMSTT	1248
ref XP_027243.1		CTGLHPPVTSYPPTLTGSHSSHSVQSGLVROSPARASVASQSS-YCYSS-RHSSLRMSTT	421
gi 15076843 gb		CTGLHPPVTSYPPTLTGSHSSHSVQSGLVROSPARASVASQSS-YCYSS-RHSSLRMSTT	2143
gi 6650377 gb A		CTGLHPPVTSYPPTLTGSHSSHSVQSGLVROSPARASVASQSS-YCYSS-RHSSLRMSTT	1248
gi 13171105 gb		-----LS--CGTSQSSQSVQSGLVRSPPARASVVSQSSSYRYSSSRHSSLRTSTT	1511
35	gi 7290294 gb A	-----LAHDRVNIGHYQLAOYHGRVAAINMCG	448
		2170 2180 2190 2200 2210 2220	
40 NOV3		GFVPCRRSSTSQISLRNLPSSIQSRLSMVNQMEESGSGSLACVQHGLPSSSSSSQSIPAC	1308
ref XP_027243.1		GFVPCRRSSTSQISLRNLPSSIQSRLSMVNQMEESGSGSLACVQHGLPSSSSSSQSIPAC	481
gi 15076843 gb		GFVPCRRSSTSQISLRNLPSSIQSRLSMVNQMEESGSGSLACVQHGLPSSSSSSQSIPAC	2203
gi 6650377 gb A		GFVPCRRSSTSQISLRNLPSSIQSRLSMVNQMEESGSGSLACVQHGLPSSSSSSQSIPAC	1308
gi 13171105 gb		GLEPCRRSSTSQISLRNLPSSIQSRLSMVNQMEESGSGSLACVQHGLPSSSSSSQSIPAC	1558
45	gi 7290294 gb A	--G-----VKKLEAVYEFFFTLIFGKG--TRYAG-----HG--SYKDVVIDGSM	485
		2230 2240 2250 2260 2270 2280	
50 NOV3		KHHTLVGFLATEGGQSSATDAQ-----PGNTLSPANNSHS--RKAEEVIYRVQIVDFSOIL	1361
ref XP_027243.1		KHHTLVGFLATEGGQSSATDAQ-----PGNTLSPANNSHS--RKAEEVIYRVQIVDFSOIL	534
gi 15076843 gb		KHHTLVGFLATEGGQSSATDAQ-----PGNTLSPANNSHS--RKAEEVIYRVQIVDFSOIL	2256
gi 6650377 gb A		KHHTLVAFILGAEGGQSSATEAC-----PGNTSSPANISHA--RKAEEVIYRVQIVDFSOIL	1361
gi 13171105 gb		KHHTLVGFLGNDGLCSTVTDPLSQHHHPHHHPQOHNPHTATVRRDDISYRVQIVDFVGOVL	1618
55	gi 7290294 gb A	EDFKFVAYFINEADTVTAASC-----G-----RDPITVAQFAELISQKCL	526
		2290 2300 2310 2320 2330 2340	
60 NOV3		EGINLSKRKELQWPDEGIRLKAGRNSWKDWSPOEGMEGHVHRWVPCSRDPGTRSHIDKA	1421
ref XP_027243.1		EGINLSKRKELQWPDEGIRLKAGRNSWKDWSPOEGMEGHVHRWVPCSRDPGTRSHIDKA	594
gi 15076843 gb		EGINLSKRKELQWPDEGIRLKAGRNSWKDWSPOEGMEGHVHRWVPCSRDPGTRSHIDKA	2316
gi 6650377 gb A		EGINLSKRKELQWPDEGIRLKAGRNSWKDWSPOEGMEGHVHRWVPCSRDPGTRSHIDNT	1421
gi 13171105 gb		ENINLSKRKELQWPDAMRHKAGRTCWDRDWSPLGMEGHVHRWVPCSRDPGTRSHIDKT	1678
65	gi 7290294 gb A	G-----RGQIEDP-----A-----TRFDWTKKLQOP-----LPQVR-----	552
		2350 2360	
70 NOV3		VLLVQIDDKYVTVIETGVLELGAEV	1446
ref XP_027243.1		VLLVQIDDKYVTVIETGVLELGAEV	619
gi 15076843 gb		VLLVQIDDKYVTVIETGVLELGAEV	2341
gi 6650377 gb A		VLLVQIDDKYVTVIETGVLELGAEV	1446
gi 13171105 gb		VLLVQVEDKIVPIIETGVLELGAEV	1703
gi 7290294 gb A		-----	552

Pecanex gene was originally discovered in *Drosophila*, encoding a large, membrane-spanning protein. The mouse homolog was recently reported. In the absence of maternal expression of the pecanex gene, the embryo develops severe hyperneuralization similar to that characteristic of Notch mutant embryos. Early gastrula embryos, lacking both maternally and zygotically expressed activity of the neurogenic pecanex locus, are shown to contain a greater than wild-type number of stably determined neural precursor cells which can differentiate into neurons in culture. Therefore it is anticipated that this novel human pecanex will be involved in neuronal differentiation, maintenance of neuronal precursors and neurological diseases.

The disclosed NOV3 nucleic acid of the invention encoding a Human homolog of the *Drosophila* pecanex protein includes the nucleic acid whose sequence is provided in Table 3A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 3A while still encoding a protein that maintains its Human homolog of the *Drosophila* pecanex activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

The disclosed NOV3 protein of the invention includes the Human homolog of the *Drosophila* pecanex protein whose sequence is provided in Table 3B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 3B while still encoding a protein that maintains its Human homolog of the *Drosophila* pecanex activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 7 percent of the residues may be so changed.

The NOV3 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in cancer, trauma, regeneration (in vitro and in vivo), viral/bacterial/parasitic infections, cardiomyopathy, atherosclerosis, hypertension, congenital

heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, multiple sclerosis, scleroderma, obesity, endometriosis, fertility, hypercoagulation, autoimmune disease, allergies, immunodeficiencies, transplantation, hemophilia, idiopathic thrombocytopenic purpura, graft versus host disease, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, systemic lupus erythematosus, asthma, emphysema, ARDS, laryngitis, psoriasis, actinic keratosis, acne, hair growth/loss, alopecia, pigmentation disorders, endocrine disorders, diabetes, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, Lesch-Nyhan syndrome, and a variety of kidney diseases and/or other pathologies and disorders.

NOV3 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV3 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated NOV3 epitope is from about amino acids 20 to 50. In another embodiment, a NOV3 epitope is from about amino acids 180 to 200. In additional embodiments, NOV3 epitopes are from about amino acids 360 to 400, from about 450 to 500, from about amino acids 600 to 680, from about amino acids 720 to 780, from about amino acids 800 to 860, from about amino acids 950 to 1000, from about amino acids 1050 to 1100, from about amino acids 1150 to 1320, and from about amino acids 1350 to 1420. These novel proteins can be used in assay systems for functional analysis of various human disorders, which are useful in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV4

A disclosed NOV4 nucleic acid of 1500 nucleotides (also referred to as SC140515441_A) encoding a novel Aurora-related kinase 1-like protein is shown in Table 4A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 182-184 and ending with a TAG codon at nucleotides 1391-1393. The start and stop codons are in bold letters, and the 5' and 3' untranslated regions are underlined.

Table 4A. NOV4 Nucleotide Sequence (SEQ ID NO:13)

```

TCATCTTAATATTTTATAGCTGATATAGTTGTAATTTCTTAACCTAGCTCATCTCTAGAGGATATGTAAAA
ACATAAAACACCTCAATTACTTGTGAATTATAGAGGTGTATCAGTTGGTTTAAAAGTGCTTTTATTGGGCT
GAGCTCTTGAAGACTCAGGTCCTTGGGTCATAGGCATCATGGACCAATCTGAAGAAAAGTGCATTTTCAGG
GCCTGTTGAGGCTAAACTCCAGTTGGAGGTCCAGAACATGTTCTCGTGAAGTCAAGCAATTTCTTGTGAGA
ATCCATTACCTGCAATAGTGGCCAGGCTCAGTGGGTCTTGTGTCCTTCAAATTTCTCGCAGCGTGTTCCT
TTGCAAGCACAAAAGCTTGTCTCCAGTCACAAGCCAGTTTCAAGATCAGAAGCAGAAGCAATTGCAGGCAAC
CAGTGTACCTCATCTGCCCTCCAGGCCACTGAATAACACCCAAAACAGCAAGCAGTCCCCGCTGTGCGCAC
CTGAAAATTAATCCTGAGGAGGAAGTGGCATCAAAACAGAAAATGAAGAATCAAAAAGAGGCAATGGGCT
TTGGAAGACCTTGAATTTGGTCGCCCTCCGGGTAAAGGAAAGTTGGTAATGTTTATTGGCAAGAGAAAA
ACAAAGCAAGTTTATTTCTGGCTCTTAGGGTGTATTATTAAGCTCAGCTGGAGAAAGCAGGAGTGGAGCATC
AACTCAGAAGAGAAGTAGAATACAGTCCACCTCCAACATCCTAATATAATCAGACTGTATGGTTATTTTC
CATGATGCCACCAGAGTCTACCTAATTTCTGGAATATACCACTTGAACAGTCAATACAGAACTTCAGAA
ACTTTCAAAGTTTGTATGAGCAGAGAACTGCTACTTATATCAGAAATGGCAAGTGCCTGTCTTACTGTC
ATTCAAAACAGTTATTCTATAGAGACATTAAGCCAGAGAACTTACTTCTTGATCAGCTGGAGAGCTTGAA
ATTGCAAAATTTTGGGTGGTCAGAACATGCTCCATCTTCCAGGAGGACCACTCTCTGTGGCACCCTGGACTA
CCTGCCCCCGAAATGATTGAAGGTGGATGCATGATGAGAAGGTGGATCTCTGGAGCCTTGGAGTTCTTT
GCTGTGAATTTTTAGTTGGGAAGCCTCCTTTTGAGGCAAAATACATACCAAGAGACCTACAAAAGAATATCA
CGGGTTGAGTTTACATTCCCTGACTTTGTAACAGAGGGAGCCAGGGACCTCATTTCAAGACTGTTGAAGCA
TGTTCCCAAGCCAGAGGCCAATGCTCAGAGAGTACTTGAATACCCCTGGATCACAGCAAAATTCATCAAAAC
CATCAAAATGCCAAAACAAAGAATCAACTAGCAAGTATTCTTAGGAATCGTGCAGGGGGAGAAATCCTTGA
GCCAGGGCTGCTGTATAACCTCTCAGGAACATGCTACCAAAATTTATTTTACCATTGACTGCTGCCCTCAA
TCTAGAACA

```

The disclosed NOV4 nucleic acid sequence maps to chromosome 1 and has 1152 of 1212 bases (95%) identical to a gb:GENBANK-ID:AF008551|acc:AF008551 mRNA from

5 *Homo sapiens* (*Homo sapiens* aurora-related kinase 1 (ARK1) mRNA, complete cds (E = 1.8e⁻²⁴³).

A disclosed NOV4 protein (SEQ ID NO:14) encoded by SEQ ID NO:13 has 403 amino acid residues, and is presented using the one-letter code in Table 4B. Signal P, Psort and/or Hydropathy results predict that NOV4 does not have a signal peptide, and is likely to be
 10 localized to the cytoplasm with a certainty of 0.4500. In other embodiments NOV4 is also likely to be localized microbody (peroxisome) with a certainty of 0.3000, to the mitochondrial membrane space with a certainty of 0.1000, or to the lysosome(lumen) with a certainty of 0.1000.

Table 4B. Encoded NOV4 protein sequence (SEQ ID NO:14).

```

MDQSEENCISGPVEAKTPVGGPEHVLVTQQFPQNPLPANSGQAQWVLCPSNSSQVRVPLQAQKLVSSHKPV
QNQKQQLQATSVHPASRPLNNTQNSKQSPLSAPENNPEELASKQKNEESKKRQWALEDLEIGRPPEGK
KEGNVYLAREKQSKFILALRVLFKAQLEKAGVEHQLRREVEIQSHLQHPNIIIRLYGYFFHDATRVYLILEYT
PLETVNTELQKLSKFDEQRTATYITELASALSCHSKTVIHRDIKPENLLGSAGELEIANFGWSEHAPSS
RRTTLCGLTDYLPPEMIEGRMHDEKVDLWSLGLVLCCEFLVGKPPFEANTYQETYKRISRVEFTFPDFVTEG
ARDLISRLLKHVPSQRPLREVLEYPWITANSSKPSNQNKESTSKYS

```

The disclosed NOV4 amino acid has 69 of 403 amino acid residues (91%) identical to, and 381 of 403 amino acid residues (94%) similar to, the 403 amino acid residue

ptnr:SPTREMBL-ACC:O60445 protein from *Homo sapiens* (Human) (Aurora-Related Kinase 1 ($E=1.7e^{-198}$)).

NOV4 is expressed in at least Adrenal Gland/Suprarenal gland, Amygdala, Bone Marrow, Brain, Cervix, Colon, Coronary Artery, Epidermis, Heart, Kidney, Liver, Lung, Lymphoid tissue, Mammary gland/Breast, Ovary, Peripheral Blood, Placenta, Prostate, Testis, Thalamus, Tonsils, Uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention.

In addition, NOV4 is predicted to be expressed in colon because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AF008551|acc:AF008551) a closely related aurora-related kinase 1 (ARK1) mRNA, complete cds homolog in species *Homo sapiens*.

NOV4 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 4C.

Table 4C. BLAST results for NOV4					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 12654873 gb AAH01280.1 AAH01280 (BC001280)	serine/threonine kinase 15 [Homo sapiens]	403	370/403 (91%)	381/403 (93%)	0.0
gi 13653970 ref XP_009546.3 (XM_009546)	serine/threonine kinase 15 [Homo sapiens]	403	369/403 (91%)	381/403 (93%)	0.0
gi 4507275 ref NP_03591.1 (NM_003600)	serine/threonine kinase 15; Serine/threonine protein kinase 15 [Homo sapiens]	403	369/403 (91%)	380/403 (93%)	0.0
gi 7446411 pir JCS974	aurora-related kinase 1 (EC 2.7.-.-) - human	403	367/403 (91%)	379/403 (93%)	0.0
gi 4507279 ref NP_03149.1 (NM_003158)	serine/threonine kinase 6; Serine/threonine protein kinase-6; serine/threonine kinase 6 (aurora/IPL1-like) [Homo sapiens]	402	342/403 (84%)	360/403 (88%)	0.0

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 4D.

Table 4D. ClustalW Analysis of NOV4

- 5 1) NOV4 (SEQ ID NO:14)
 2) gi|12654873|gb|AAH01280.1|AAH01280 (BC001280) serine/threonine kinase 15 [Homo sapiens] (SEQ ID NO:47)
 3) gi|13653970|ref|XP_009546.3| (XM_009546) serine/threonine kinase 15 [Homo sapiens] (SEQ ID NO:48)
 10 4) gi|4507275|ref|NP_003591.1| (NM_003600) serine/threonine kinase 15; Serine/threonine protein kinase 15 [Homo sapiens] (SEQ ID NO:49)
 5) gi|7446411|pir|JC5974 aurora-related kinase 1 (EC 2.7.-.-) - human (SEQ ID NO:50)
 15 6) gi|4507279|ref|NP_003149.1| (NM_003158) serine/threonine kinase 6; Serine/threonine protein kinase-6; serine/threonine kinase 6 (aurora/IPL1-like) [Homo sapiens] (SEQ ID NO:51)

		10	20	30	40	50		
20	NOV4	MDQSEENCISGPV	EAKTIPVGGPEH	VLVTQQFP	CONPLPANSGQAQV	LCP	
	gi 12654873	MDRSKENCISGPV	KATAPVGGPKRV	LVLTQQFP	CONPLPVNSGQAQV	LCP	
	gi 13653970	MDRSKENCISGPV	KATAPVGGPKRV	LVLTQQFP	CONPLPVNSGQAQV	LCP	
	gi 4507275	MDRSKENCISGPV	KATAPVGGPKRV	LVLTQQFP	CONPLPVNSGQAQV	LCP	
25	gi 7446411	MDRSKENCISGPV	KATAPVGGPKRV	LVLTQQFP	CONPLPVNSGQAQV	LCP	
	gi 4507279	MDRSKENCISGPV	KATAPVGGPKRV	LVLTQQFP	CONPLPVNSGQAQV	LCP	
		60	70	80	90	100		
30	NOV4	SNSSQVRVPLQAQ	KLVS	SHKPVQ	NQKQKQLQATSV	PHPASRPLNNTQ	NSKQ
	gi 12654873	SNSSQVRVPLQAQ	KLVS	SHKPVQ	NQKQKQLQATSV	PHPVSRPLNNTQ	SKSQ
	gi 13653970	SNSSQVRVPLQAQ	KLVS	SHKPVQ	NQKQKQLQATSV	PHPVSRPLNNTQ	SKSQ
	gi 4507275	SNSSQVRVPLQAQ	KLVS	SHKPVQ	NQKQKQLQATSV	PHPVSRPLNNTQ	SKSQ
	gi 7446411	SNSSQVRVPLQAQ	KLVS	SHKPVQ	NQKQKQLQATSV	PHPVSRPLNNTQ	SKSQ
35	gi 4507279	SNSSQVRVPLQAQ	KLVS	SHKPVQ	NQKQKQLQATSV	PHPVSRPLNNTQ	SKSQ
		110	120	130	140	150		
40	NOV4	SPLSAPENNPEEEL	ASKQKNEESK	RQWALEDFEIGR	PLGKGF	GNVYLA	
	gi 12654873	PLPSAPENNPEEEL	ASKQKNEESK	RQWALEDFEIGR	PLGKGF	GNVYLA	
	gi 13653970	PLPSAPENNPEEEL	ASKQKNEESK	RQWALEDFEIGR	PLGKGF	GNVYLA	
	gi 4507275	PLPSAPENNPEEEL	ASKQKNEESK	RQWALEDFEIGR	PLGKGF	GNVYLA	
	gi 7446411	PLPSAPENNPEEEL	ASKQKNEESK	RQWALEDFEIGR	PLGKGF	GNVYLA	
45	gi 4507279	PLPSAPENNPEEEL	ASKQKNEESK	RQWALEDFEIGR	PLGKGF	GNVYLA	
		160	170	180	190	200		
50	NOV4	REKQSKFILALR	VLFAQLEKAG	VEHQLRREVEIQ	SHLRHPN	ILRLYGYF	
	gi 12654873	REKQSKFILALK	VLFKAQLEKAG	VEHQLRREVEIQ	SHLRHPN	ILRLYGYF	
	gi 13653970	REKQSKFILALK	VLFKAQLEKAG	VEHQLRREVEIQ	SHLRHPN	ILRLYGYF	
	gi 4507275	REKQSKFILALK	VLFKAQLEKAG	VEHQLRREVEIQ	SHLRHPN	ILRLYGYF	
	gi 7446411	REKQSKFILALK	VLFKAQLEKAG	VEHQLRREVEIQ	SHLRHPN	ILRLYGYF	
	gi 4507279	REKQSKFILALK	VLFKAQLEKAG	VEHQLRREVEIQ	SHLRHPN	ILRLYGYF	
55		210	220	230	240	250		
	NOV4	HDATRVYLILEY	APLGT	VYRELQKLSK	FDEQRTATYITEL	ANALSYCHSK	
	gi 12654873	HDATRVYLILEY	APLGT	VYRELQKLSK	FDEQRTATYITEL	ANALSYCHSK	
	gi 13653970	HDATRVYLILEY	APLGT	VYRELQKLSK	FDEQRTATYITEL	ANALSYCHSK	
60	gi 4507275	HDATRVYLILEY	APLGT	VYRELQKLSK	FDEQRTATYITEL	ANALSYCHSK	
	gi 7446411	HDATRVYLILEY	APLGT	VYRELQKLSK	FDEQRTATYITEL	ANALSYCHSK	
	gi 4507279	HDATRVYLILEY	APLGT	VYRELQKLSK	FDEQRTATYITEL	ANALSYCHSK	
		260	270	280	290	300		
65	NOV4	TVIHRDIKPENLL	GSAGELEIAN	FGWSEHAPSSR	RTTLCGTL	DYLPPEM	

5	gi 12654873	RVIHRDIKPENLLLSAGELKIADFGWSVHAPSSRRTTLCGTLDYLPPEM
	gi 13653970	RVIHRDIKPENLLLSAGELKIADFGWSVHAPSSRRTTLCGTLDYLPPEM
	gi 4507275	RVIHRDIKPENLLLSAGELKIADFGWSVHAPSSRRTTLCGTLDYLPPEM
	gi 7446411	RVIHRDIKPENLLLSAGELKIADFGWSVHAPSSRRTTLCGTLDYLPPEM
	gi 4507279	RVIHRDIKPENLLLSAGELKIADFGWSVHAPSSRRTTLCGTLDYLPPEM
<div> <div>310320330340350</div> <div> <div>..... </div> <div> <div>NOV4</div> <div>IEGRMHDEKVDLWSLGVLCCEFLVGKPPFEANTYQETYKRISRVEFTFPD</div> <div>gi 12654873 </div> <div>IEGRMHDEKVDLWSLGVLCYEFLVGKPPFEANTYQETYKRISRVEFTFPD</div> <div>gi 13653970 </div> <div>IEGRMHDEKVDLWSLGVLCYEFLVGKPPFEANTYQETYKRISRVEFTFPD</div> <div>gi 4507275 </div> <div>IEGRMHDEKVDLWSLGVLCYEFLVGKPPFEANTYQETYKRISRVEFTFPD</div> <div>gi 7446411 </div> <div>IEGRMHDEKVDLWSLGVLCYEFLVGKPPFEANTYQETYKRISRVEFTFPD</div> <div>gi 4507279 </div> <div>IEGRMHDEKVDLWSLGVLCYEFLVGKPPFEANTYQETYKRISRVEFTFPD</div> </div> </div> </div>		
10	gi 12654873	IEGRMHDEKVDLWSLGVLCYEFLVGKPPFEANTYQETYKRISRVEFTFPD
	gi 13653970	IEGRMHDEKVDLWSLGVLCYEFLVGKPPFEANTYQETYKRISRVEFTFPD
	gi 4507275	IEGRMHDEKVDLWSLGVLCYEFLVGKPPFEANTYQETYKRISRVEFTFPD
	gi 7446411	IEGRMHDEKVDLWSLGVLCYEFLVGKPPFEANTYQETYKRISRVEFTFPD
	gi 4507279	IEGRMHDEKVDLWSLGVLCYEFLVGKPPFEANTYQETYKRISRVEFTFPD
<div> <div>360370380390400</div> <div> <div>..... </div> <div> <div>NOV4</div> <div>FVTEGARDLISRLKHNPSQRPMLREVLEHPWITANSSKPSNCQNKESAS</div> <div>gi 12654873 </div> <div>FVTEGARDLISRLKHNPSQRPMLREVLEHPWITANSSKPSNCQNKESAS</div> <div>gi 13653970 </div> <div>FVTEGARDLISRLKHNPSQRPMLREVLEHPWITANSSKPSNCQNKESAS</div> <div>gi 4507275 </div> <div>FVTEGARDLISRLKHNPSQRPMLREVLEHPWITANSSKPSNCQNKESAS</div> <div>gi 7446411 </div> <div>FVTEGARDLISRLKHNPSQRPMLREVLEHPWITANSSKPSNCQNKESAS</div> <div>gi 4507279 </div> <div>FVTEGARDLISRLKHNPSQRPMLREVLEHPWITANSSKPSNCQNKESAS</div> </div> </div> </div>		
20	gi 12654873	FVTEGARDLISRLKHNPSQRPMLREVLEHPWITANSSKPSNCQNKESAS
	gi 13653970	FVTEGARDLISRLKHNPSQRPMLREVLEHPWITANSSKPSNCQNKESAS
	gi 4507275	FVTEGARDLISRLKHNPSQRPMLREVLEHPWITANSSKPSNCQNKESAS
	gi 7446411	FVTEGARDLISRLKHNPSQRPMLREVLEHPWITANSSKPSNCQNKESAS
	gi 4507279	FVTEGARDLISRLKHNPSQRPMLREVLEHPWITANSSKPSNCQNKESAS
<div> <div>...</div> <div> <div>NOV4</div> <div>KYS</div> <div>gi 12654873 </div> <div>KQS</div> <div>gi 13653970 </div> <div>KQS</div> <div>gi 4507275 </div> <div>KQS</div> <div>gi 7446411 </div> <div>KQS</div> <div>gi 4507279 </div> <div>KQS</div> </div> </div>		
30	gi 12654873	KQS
	gi 13653970	KQS
	gi 4507275	KQS
	gi 7446411	KQS
	gi 4507279	KQS

Tables 4E-G lists the domain description from DOMAIN analysis results against NOV4. This indicates that the NOV4 sequence has properties similar to those of other proteins known to contain this domain.

Table 4E Domain Analysis of NOV4

gnl|Smart|smart00220, S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. (SEQ ID NO:98)
 CD-Length = 256 residues, 99.6% aligned
 Score = 256 bits (653), Expect = 2e-69

40

NOV 3:	134	EIGRPPGKGGKFGNVYLAREKQSKFILALRVLFKAQLEKAGVEHQLRREVEIQSHLQHPNI	193
		+ + ++ ++ ++ + + + ++ ++	
Sbjct:	2	ELLEVLGKGAFGKVYLARDKKTGLVAIKVIKKEKLKK-KKRERILREIKILKKLDHPNI	60
NOV 3:	194	IRLYGYFHDA TRVYLILEYTPLETVNTELQKLSKFDEQRTATYITELASALSYCHSKTVI	253
		++ ++ ++ + + + ++ + +	
Sbjct:	61	VKLYDVVFEDDDKLYLVMEYCEGGDLFDLLKKRGLSEDEARFYARQILSALEYLHSGQII	120
NOV 3:	254	HRDIKPENLLLSAGELEIANFGWS--EHAPSSRRTTLCGTLDYLPPEMIEGRMHDEKVD	311
		+ + ++ ++ + + + + ++ + + +	
Sbjct:	121	HRDLKPENILLDSGHVKLADFGGLAQQLDGGTLLTTFVGTPEYMAPEVLLGKGYGKAVD	180
NOV 3:	312	LWSLGVLCCEFLVGKPPFEA-NTYQETYKRISRVEFTFPDF---VTEGARDLISRLKHN	367
		+ ++ + + + + ++ ++ +	
Sbjct:	181	IWSLGVILYELLTGKPPFPDQQLALFKKIGKPPPPPPPEWKISPEAKDLIKKLLVKD	240

5

gnl|Pfam|pfam00069, kinase, Protein kinase domain (SEQ ID NO:99)
CD-Length = 256 residues, 100.0% aligned
Score = 221 bits (564), Expect = 5e-59

10

15

20

25

gnl|Smart|smart00219, TyrKc, Tyrosine kinase, catalytic domain;
Phosphotransferases. Tyrosine-specific kinase subfamily (SEQ ID
NO:100)
CD-Length = 258 residues, 99.6% aligned
Score = 127 bits (318), Expect = 2e-30

30

35

40

45

Amplification of chromosome 20q DNA has been reported in a variety of cancers. DNA amplification on 20q13 has also been correlated with poor prognosis among axillary node-negative breast tumor cases. Sen et al. (1997) cloned a partial cDNA encoding STK15 (also known as BTAK and aurora2) from this amplicon and found that it is amplified and overexpressed in 3 human breast cancer cell lines. STK15 encodes a centrosome-associated kinase. Zhou et al. (1998) found that STK15 is involved in the induction of centrosome duplication-distribution abnormalities and aneuploidy in mammalian cells. Centrosomes appear to maintain genomic stability through the establishment of bipolar spindles during cell division, ensuring equal segregation of replicated chromosomes to 2 daughter cells. Deregulated duplication and distribution of centrosomes are implicated in chromosome segregation abnormalities, leading to aneuploidy seen in many cancer cell types. Zhou et al. (1998) found amplification of STK15 in approximately 12% of primary breast tumors, as well as in breast, ovarian, colon, prostate, neuroblastoma, and cervical cancer cell lines. Additionally, high expression of STK15 mRNA was detected in tumor cell lines without evidence of gene amplification. Ectopic expression of STK15 in mouse NIH 3T3 cells led to the appearance of abnormal centrosome number (amplification) and transformation in vitro. Finally, overexpression of STK15 in near-diploid human breast epithelial cells revealed similar centrosome abnormality, as well as induction of aneuploidy. These findings suggested that STK15 is a critical kinase-encoding gene, whose overexpression leads to centrosome amplification, chromosomal instability, and transformation in mammalian cells. Zhou et al. (1998) found that the open reading frame of the full-length STK15 cDNA sequence encodes a 403-amino acid protein with a molecular mass of approximately 46 kD. STK6 (602687), also referred to as AIK, is highly homologous to STK15. The *Drosophila* 'aurora' and *S. cerevisiae* Ipl1 STKs are involved in mitotic events such as centrosome separation and chromosome segregation. Using a degenerate primer-based PCR method to screen for novel STKs, Shindo et al. (1998) isolated mouse and human cDNAs encoding STK15, which they termed ARK1 (aurora-related kinase-1). Cell cycle and Northern blot analyses showed that peak expression of STK15 occurs during the G2/M phase and then decreases. By interspecific backcross mapping, Shindo et al. (1998) mapped the mouse *Stk15* gene to the distal region of chromosome 2 in a region showing homology of synteny with human 20q

The disclosed NOV4 nucleic acid of the invention encoding a Aurora-related kinase 1-like protein includes the nucleic acid whose sequence is provided in Table 4A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 4A while still encoding a protein that

maintains its Aurora-related kinase 1-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes
5 nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense
10 binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 5 percent of the bases may be so changed.

The disclosed NOV4 protein of the invention includes the Aurora-related kinase 1-like protein whose sequence is provided in Table 4B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown
15 in Table 4B while still encoding a protein that maintains its Aurora-related kinase 1-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 9 percent of the residues may be so changed.

The protein similarity information, expression pattern, and map location for the Aurora-related kinase 1-like protein and nucleic acid (NOV4) disclosed herein suggest that
20 NOV4 may have important structural and/or physiological functions characteristic of the citron kinase-like family. Therefore, the NOV4 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic
25 applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo.

The NOV4 nucleic acids and proteins of the invention are useful in potential diagnostic
30 and therapeutic applications implicated in various diseases and disorders described below. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from breast, ovarian, colon, prostate, neuroblastoma, and cervical cancer, Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary

stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberosclerosis, Scleroderma, Obesity, Transplantation, Diabetes, Von Hippel-Lindau (VHL) syndrome, Pancreatitis, Alzheimer's disease, Stroke, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis,

5. Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, and Neuroprotection and/or other pathologies. The NOV4 nucleic acid, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV4 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV4 protein have multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, contemplated NOV4 epitope is from about amino acids 1 to 10. In another embodiment, a NOV4 epitope is from about amino acids 15 to 160. In additional embodiments, NOV4 epitopes are from about amino acids 175 to 210, from about amino acids 220 to 240, from about amino acids 250 to 270, from about amino acids 280 to 320, from about amino acids 340 to 375, and from about amino acids 380 to 400. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV5

A disclosed NOV5 nucleic acid of 1500 nucleotides (designated CuraGen Acc. No. SC44326718_A) encoding a novel 26S protease regulatory subunit 4-like protein is shown in Table 5A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 101-103 and ending with a TAG codon at nucleotides 1427-1429. A putative untranslated region downstream from the termination codon is underlined in Table 5A, and the start and stop codons are in bold letters.

Table 5A. NOV5 Nucleotide Sequence (SEQ ID NO:15)

```

GTATCCCCAAGAGAAAATACGCATCAAAAATTAGGAACCTAGAAATGATAGTTGAGGTGGAGGAACTTCC
AGCAGTGGCAGCTCAAGTGGCCAAGACAAGATGGGTCAAAGTCAGGGTGATGGTCATGGTCCTAGACGTG
GCAAGAAGGATGAAAAGGACAAGAAAAATAAGTACGAACCTCTTGTACCAACTAGAGTGGCGGAAAAAGA
AGAAAAACAAAGGGACAAGATGTTGCCAGTAACTGCCACTGGTGACACTTCACACTCAGTGTCTCGGTTA
AAATTACTGAAGTTAGAGAGAATTAAAGACTACCTTCTCATGGTGAAGAATTATTAGAAATCAGGAAC
AAATAAACTATTAGAAGAAAAGCAAGAGGAGGGAAGATCAAAGTGGATGATCTGAGGGGACCCCAAT

```

```

GTCAGTAGGAAACTTGGAGAGATCATCGATGACAATCATGCCATTGTGTCTACATCTGTGGGCTCAGAA
CACTATGACAGCATTATTTTCATTGTAGAGAAGGATCTGCTGGAACCTGGCTGCTCGATTCTGCTCAGAC
ACAAGGTACATGCGGTGATAGGGGTGCTGATGGATGATACGGGTCCCCTGGTCACAATGATGAAGGTGGA
GAAGGCCCCCAGGAGACCTATGTCAATACTGGGGGTGGACAACCAAATTGAGAAATTAAGGAATCT
ATGGAGCTTCCTCTCCCCATCTGAATATTATGAAGAGATGGGTACAAAGCCTCCTAAAGGGGTCAATTC
TCTGTGGTCCACTGGCAGAGTAAACCTTGTAGCCAAAGCAGTAGCAAAACCAACCTCAGCCACTTT
CTTGAGAGTGGTGGCTCTGAACCTATTGAGAGTACCTAGGTGATGGGCCCCAACTCGTACGGCAAGTA
TTTCAAGTTGCTGAAGAATGCACCATCCATCATGTTTACTGATGAAATGAAGCCATTGGGACAAAAA
GATATGACTCCAATTCTGGTGGTGAGAGAGAAATTGAGCAACAATGTTGGAATTGGAACGTTGAACCA
ATTGGGTGGATTGATTCTAGGGAAGATGTGAAGTTATCATGGCCACAAAAAAGTAGAAACTTTGGAT
CCAGTACTTATCAGACCAGGCCGATGACAAGAAGATCGAGTCCACCTGCCTGATGAAAAGACTAAGA
AGCACATCTTTTTCAGATTACACAGCAGGATGACACTGGCCAATGATGTAACCTGGACGACTTGATCAT
GGCTAAAGATGACTTCTCTGGTCTGACATCAAGGCAATCTGTACAGAAGCTGGTCTGATGGCCTTAAGA
GAACATAGAATGAAAGCAACAAATGAAGACTTCAAAAAATCTATAGAAAGTGTCTTTATAAGAAACACG
AAGGCATCCCTGAGGGGCTTTATCTCTAGTGAACCACCGCTGCCATCAGGAAGATGGTTGGGAGATTTC
CAACCCCTGAAAGGGATGAGGTTGGGGGAG

```

The nucleic acid sequence NOV5, located on chromosome 5 has 1347 of 1447 bases (93%) identical to a gb:GENBANK-ID:HUM26SPSIV|acc:L02426 mRNA from *Homo sapiens* (Human 26S protease (S4) regulatory subunit mRNA, complete cds ($E = 2.4e^{-277}$)).

A NOV5 polypeptide (SEQ ID NO:16) encoded by SEQ ID NO:15 is 442 amino acid residues and is presented using the one letter code in Table 5B. Signal P, Psort and/or Hydropathy results predict that NOV5 has no signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.4500. In other embodiments, NOV5 may also be localized to the microbody (peroxisome) with a certainty of 0.3000, the mitochondrial matrix space with a certainty of 0.1000, or the lysosome (lumen) with a certainty of 0.1000.

Table 5B. NOV5 protein sequence (SEQ ID NO:16)

```

MGQSQGDGHGPRRGKKDEKDKKNKYELVPTRVAEKEEKTGQDVASKLPLVTLHTQCRLKLLKLERIKDYLLM
VEEFIRNQEQIKLLEEKQEEGRSKVDDLRTGTPMSVGNLEEIIIDNHAIVSTSVGSEHYDSIISFVEKDILLEPGC
SILLRHKVHAVIGVLMDDTGPLVTMMKVEKAPQETVNTGGLDNQIQEIKESMELPLPHPEYYEEMGTPPKGV
ILCGPPGTGKTLAKAVANQTSATFLRVVGSELIQKYLGDGPKLVQRQVQVAEEHAPSIMFTDEIEAIGTKRYD
SNSGGEREIQQTMLELELLNLGGFDSREDVKVIMATKQVETLDPVLIRPGRIDKKIEFHLPEDETKKHIFQIH
TSRMTLANDVTLDDLIMAKDDFSGADIKICTEAGLMALREHRMKATNEDFKKSIESVLYKKHEGIPEGLYL

```

The full amino acid sequence of the protein of the invention was found to have 383 of 442 amino acid residues (86%) identical to, and 405 of 442 amino acid residues (91%) similar to, the 440 amino acid residue ptnr:SWISSPROT-ACC:P49014 protein from *Mus musculus* (Mouse), and *Rattus norvegicus* (Rat) (26S Protease Regulatory Subunit 4 (P26S4) ($E = 1.7e^{-200}$)).

NOV5 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 5C.

Table 5C. BLAST results for NOV5

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 4506207 ref NP_02793.1 (NM_002802)	proteasome (prosome, macropain) 26S subunit, ATPase, 1; Proteasome 26S subunit, ATPase, 1 [Homo sapiens]	440	382/442 (86%)	405/442 (91%)	0.0
gi 6679501 ref NP_032973.1 (NM_008947)	protease (prosome, macropain) 26S subunit, ATPase 1 [Mus musculus]	440	383/442 (86%)	405/442 (90%)	0.0
gi 345717 pir A44468	26S proteasome regulatory chain 4 [validated] - human	440	381/442 (86%)	404/442 (91%)	0.0
gi 16741033 gb AAH16368.1 AAH16368 (BC016368)	protease (prosome, macropain) 26S subunit, ATPase 1 [Homo sapiens]	440	382/442 (86%)	404/442 (90%)	0.0
gi 2492516 sp Q90732 PRS4_CHICK	26S PROTEASE REGULATORY SUBUNIT 4 (P26S4)	440	378/442 (85%)	402/442 (90%)	0.0

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 5D.

Table 5D ClustalW Analysis of NOV5

- 1) NOV5 (SEQ ID NO:16)
- 2) gi|4506207|ref|NP_002793.1| (NM_002802) proteasome (prosome, macropain) 26S subunit, ATPase, 1; Proteasome 26S subunit, ATPase, 1 [Homo sapiens] (SEQ ID NO:52)
- 3) gi|6679501|ref|NP_032973.1| (NM_008947) protease (prosome, macropain) 26S subunit, ATPase 1 [Mus musculus] (SEQ ID NO:53)
- 4) gi|345717|pir||A44468 26S proteasome regulatory chain 4 [validated] - human (SEQ ID NO:54)
- 5) gi|16741033|gb|AAH16368.1|AAH16368 (BC016368) protease (prosome, macropain) 26S subunit, ATPase 1 [Homo sapiens] (SEQ ID NO:55)
- 6) gi|2492516|sp|Q90732|PRS4_CHICK 26S PROTEASE REGULATORY SUBUNIT 4 (P26S4) (SEQ ID NO:56)

		10	20	30	40	50
	NOV5	MGQSQGDGHGFRRGKKDEKDKKNKYEPVLPTRVAEKKEKTKGQDVASKLP			
20	gi 4506207	MGQSQSGGHGPGGGKKDDEKDKKKYEPPVPTRVGKKKKKTKGPDAAASKLP			
	gi 6679501	MGQSQSGGHGPGGGKKDDEKDKKKYEPPVPTRVGKKKKKTKGPDAAASKLP			
	gi 345717	MGQSQSGGHGPGGGKKDDEKDKKKYEPPVPTRVGKKKKKTKGPDAAASKLP			
	gi 16741033	MGQSQSGGHGPGGGKKDDEKDKKKYEPPVPTRVGKKKKKTKGPDAAASKLP			
25	gi 2492516	MGQSQSGGHGPGGGKKDDEKDKKKYEPPVPTRVGKKKKKTKGPDAAASKLP			
		60	70	80	90	100
	NOV5	LVTLPHTQCRLLKLERIKDYLLMVEEFIRNQEQIKLLEEKQEEGRSKVD			
30	gi 4506207	LVTLPHTQCRLLKLERIKDYLLMVEEFIRNQEQMKPLEEKQEEERSKVD			
	gi 6679501	LVTLPHTQCRLLKLERIKDYLLMVEEFIRNQEQMKPLEEKQEEERSKVD			
	gi 345717	LVTLPHTQCRLLKLERIKDYLLMVEEFIRNQEQMKPLEEKQEEERSKVD			
	gi 16741033	LVTLPHTQCRLLKLERIKDYLLMVEEFIRNQEQMKPLEEKQEEERSKVD			
	gi 2492516	LVTLPHTQCRLLKLERIKDYLLMVEEFIRNQEQMKPLEEKQEEERSKVD			
35		110	120	130	140	150
	NOV5	DLRGTFMSVGNLEEIIDDNHAIVSTSVGSEHYDSISFVSKOLLEPGCSV			
	gi 4506207	DLRGTFMSVGTLEEIIDDNHAIVSTSVGSEHYVSILSFVDKOLLEPGCSV			

5	gi 6679501 gi 345717 gi 16741033 gi 2492516	DLRGT PM SVGTLEEIIIDNHAIVSTSVGSEHYVSILSFVDKDLLEPGCSV DLRGT PM SVGTLEEIIIDNHAIVSTSVGSEHYVSILSFVDKDLLEPGCSV DLRGT PM SVGTLEEIIIDNHAIVSTSVGSEHYVSILSFVDKDLLEPGCSV DLRGT PM SVGTLEEIIIDNHAIVSTSVGSEHYVSILSFVDKDLLEPGCSV
10	NOV5 gi 4506207 gi 6679501 gi 345717 gi 16741033 gi 2492516	<div>160 170 180 190 200</div> <div>..... </div> LLNHK VHAVIGVLMDDT GL PLVT MM KVEKAPQETY VNT GGLDNQIQEIKES LLNHK VHAVIGVLMDDT GL PLVT MM KVEKAPQETYADIGGLDNQIQEIKES LLNHK VHAVIGVLMDDT GL PLVT MM KVEKAPQETYADIGGLDNQIQEIKES LLNHK VHAVIGVLMDDT GL PLVT MM KVEKAPQETYADIGGLDNQIQEIKES LLNHK VHAVIGVLMDDT GL PLVT MM KVEKAPQETYADIGGLDNQIQEIKES LLNHK VHAVIGVLMDDT GL PLVT MM KVEKAPQETYADIGGLDNQIQEIKES
15	NOV5 gi 4506207 gi 6679501 gi 345717 gi 16741033 gi 2492516	<div>210 220 230 240 250</div> <div>..... </div> MELPL THPEYYEEMG IK PPKGVIL CG PPPGTGKTLAKAVANQTSATFLRV MELPL THPEYYEEMG IK PPKGVILY CG PPPGTGKTLAKAVANQTSATFLRV MELPL THPEYYEEMG IK PPKGVILY CG PPPGTGKTLAKAVANQTSATFLRV MELPL THPEYYEEMG IK PPKGVILY CG PPPGTGKTLAKAVANQTSATFLRV MELPL THPEYYEEMG IK PPKGVILY CG PPPGTGKTLAKAVANQTSATFLRV MELPL THPEYYEEMG IK PPKGVILY CG PPPGTGKTLAKAVANQTSATFLRV
20	NOV5 gi 4506207 gi 6679501 gi 345717 gi 16741033 gi 2492516	<div>260 270 280 290 300</div> <div>..... </div> VGSELIQ YLG DP GPKLV RV QV AE EHAPSIVFIDEIDAIGTKRYDSNSG VGSELIQ YLG DP GPKLV REL FRVAEEHAPSIVFIDEIDAIGTKRYDSNSG VGSELIQ YLG DP GPKLV REL FRVAEEHAPSIVFIDEIDAIGTKRYDSNSG VGSELIQ YLG DP GPKLV REL FRVAEEHAPSIVFIDEIDAIGTKRYDSNSG VGSELIQ YLG DP GPKLV REL FRVAEEHAPSIVFIDEIDAIGTKRYDSNSG VGSELIQ YLG DP GPKLV REL FRVAEEHAPSIVFIDEIDAIGTKRYDSNSG
25	NOV5 gi 4506207 gi 6679501 gi 345717 gi 16741033 gi 2492516	<div>310 320 330 340 350</div> <div>..... </div> GEREIQ RTM LE LLNQLDGFDSRGDVKVIMATNRIETLDPALIRPGRID GEREIQ RTM LE LLNQLDGFDSRGDVKVIMATNRIETLDPALIRPGRID GEREIQ RTM LE LLNQLDGFDSRGDVKVIMATNRIETLDPALIRPGRID GEREIQ RTM LE LLNQLDGFDSRGDVKVIMATNRIETLDPALIRPGRID GEREIQ RTM LE LLNQLDGFDSRGDVKVIMATNRIETLDPALIRPGRID GEREIQ RTM LE LLNQLDGFDSRGDVKVIMATNRIETLDPALIRPGRID
30	NOV5 gi 4506207 gi 6679501 gi 345717 gi 16741033 gi 2492516	<div>360 370 380 390 400</div> <div>..... </div> RKIEF PLPDEKTKKRI FQ IHTSRMTLAD VT LDLIMAKDDL SG ADIKAI RKIEF PLPDEKTKKRI FQ IHTSRMTLAD VT LDLIMAKDDL SG ADIKAI RKIEF PLPDEKTKKRI FQ IHTSRMTLAD VT LDLIMAKDDL SG ADIKAI RKIEF PLPDEKTKKRI FQ IHTSRMTLAD VT LDLIMAKDDL SG ADIKAI RKIEF PLPDEKTKKRI FQ IHTSRMTLAD VT LDLIMAKDDL SG ADIKAI RKIEF PLPDEKTKKRI FQ IHTSRMTLAD VT LDLIMAKDDL SG ADIKAI
35	NOV5 gi 4506207 gi 6679501 gi 345717 gi 16741033 gi 2492516	<div>410 420 430 440</div> <div>..... </div> CTEAGL MALR RR RMKVTNEDFKKSKENVLYKKQEGTPEGLYL CTEAGL MALR RR RMKVTNEDFKKSKENVLYKKQEGTPEGLYL CTEAGL MALR RR RMKVTNEDFKKSKENVLYKKQEGTPEGLYL CTEAGL MALR RR RMKVTNEDFKKSKENVLYKKQEGTPEGLYL CTEAGL MALR RR RMKVTNEDFKKSKENVLYKKQEGTPEGLYL CTEAGL MALR RR RMKVTNEDFKKSKENVLYKKQEGTPEGLYL

Tables 5E-F list the domain description from DOMAIN analysis results against NOV5. This indicates that the NOV5 sequence has properties similar to those of other proteins known to contain this domain.

Table 5E. Domain Analysis of NOV5

gnl|Pfam|pfam00004, AAA, ATPase family associated with various cellular activities (AAA). AAA family proteins often perform chaperone-like functions that assist in the assembly, operation, or disassembly of protein complexes (SEQ ID NO:101)
 CD-Length = 186 residues, 100.0% aligned
 Score = 190 bits (483), Expect = 1e-49

NOV 4:	221	GVILCGPPGTGKTLAKAVANQTSATFLRVVGSSELIQKYLGDGPKLVRQVFQVAEEHAPS	280
		++ + + + + + + +	
Sbjct:	1	GILLYGPPGTGKTLAKAVAKELGVPFIEISGSELLSKYVGSEKLVRAFLSLARKSAPC	60
NOV 4:	281	IMFTDEIEAIGTKRYDSNSGGEREIQOTMLELELLNQLGGFDSREDVKVIMATKQVETLD	340
		+ + + + + + ++ + + + +	
Sbjct:	61	IIFIDEIDALAPKRGDVGTDVSS---RVVNQLLEMDGFEKLSNVIVIGATNRPDLLD	116
NOV 4:	341	PVLIRPGRIDKKIEFHLPEDEKTKKHIFQIHTSRMTLANDVTLDLIMAKDDFSGADIKAI	400
		+ ++ + + + + ++ + +	
Sbjct:	117	PALLRPGRFDRRIEVLPEDEERLEILKHLKKKPLEKVDLDEIARRTPGFSGADLAAL	176
NOV 4:	401	CTEAGLMALR	410
		+	
Sbjct:	177	CREAALRAIR	186

Table 5F. Domain Analysis of NOV5

gnl|Smart|smart00382, AAA, ATPases associated with a variety of cellular activities; AAA. This profile/alignment only detects a fraction of this vast family. The poorly conserved N-terminal helix is missing from the alignment. (SEQ ID NO:102)
 CD-Length = 151 residues, 100.0% aligned
 Score = 61.6 bits (148), Expect = 9e-11

NOV 4:	218	PKGVILCGPPGTGKTLAKAVANQTSATFLRVV-----GSELIQK	258
		+ ++ + + + +	
Sbjct:	1	PGEVVLIVGPPGSGKTTIARALARELGPDGGGVIIYIDGEDLREEALLQLRLLLVLVGEDK	60
NOV 4:	259	YLGDPKLVLRQVFQVAEEHAPSIMFTDEIEAIGTKRYDSNSGGEREIQOTMLELELLNQL	318
		+ + + + ++ ++ + +	
Sbjct:	61	LSGSGGQRIRLALALARKLKPVDVILDEITSLLDAEQE-----ALLLLEELLRL	111
NOV 4:	319	GGFDSREDVKVIMATKQVETLDPVLIRPGRIDKKIEFHLPD	359
		+ + ++	
Sbjct:	112	LLLLKEENVTVIATTNDETDLIPALLRR-RFDRRIVLLRIL	151

Ubiquitinated proteins are degraded by a 26S ATP-dependent protease. The protease is composed of a 20S catalytic proteasome and 2 PA700 regulatory modules. The PA700 complex is composed of multiple subunits, including at least 6 related ATPases and approximately 15 non-ATPase polypeptides. Tanahashi et al. (1998) stated that each of the 6 ATPases, namely PSMC1, PSMC2 (154365), PSMC3 (186852), PSMC4 (602707), PSMC5 (601681), and PSMC6 (602708), contains an AAA (ATPases associated with diverse cellular activities) domain (see PSMC5). Dubiel et al. (1992) cloned cDNAs encoding subunit 4 (S4) of the 26S protease by screening a HeLa cell cDNA library with probes that were produced

using the protein sequence. The 440-amino acid protein has a molecular mass of 51 kD by SDS-PAGE. By fluorescence in situ hybridization, Tanahashi et al. (1998) mapped the human PSMC1 gene to 19p13.3. Hoyle and Fisher (1996) found that the human and mouse PSMC1 proteins have 99% amino acid identity. They reported that the mouse Psmc1 gene contains at least 11 exons. By analysis of an interspecific backcross, Hoyle and Fisher (1996) mapped the mouse Psmc1 gene to chromosome 12. Nomenclature note: The PSMC1 gene product, which Dubiel et al. (1992) called subunit 4 (S4), is distinct from the PSMC4 (602707) gene product.

The disclosed NOV5 nucleic acid of the invention encoding a 26S protease regulatory subunit 4 -like protein includes the nucleic acid whose sequence is provided in Table 5A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 5A while still encoding a protein that maintains its 26S protease regulatory subunit 4 -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 7 percent of the bases may be so changed.

The disclosed NOV5 protein of the invention includes the 26S protease regulatory subunit 4 -like protein whose sequence is provided in Table 5B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 5B while still encoding a protein that maintains its 26S protease regulatory subunit 4 -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 14 percent of the residues may be so changed.

The protein similarity information, expression pattern, and map location for the 26S protease regulatory subunit 4-like protein and nucleic acid (NOV5) disclosed herein suggest that this NOV5 protein may have important structural and/or physiological functions characteristic of the 26S protease regulatory subunit 4 family. Therefore, the NOV5 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic

applications. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, 5 diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo.

The NOV5 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below. For 10 example, the compositions of the present invention will have efficacy for treatment of patients suffering from cataract and Aphakia, Alzheimer's disease, neurodegenerative disorders, inflammation and modulation of the immune response, viral pathogenesis, aging-related disorders, neurologic disorders, cancer and/or other pathologies. The NOV5 nucleic acids, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or 15 amount of the nucleic acid or the protein are to be assessed.

NOV5 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" 20 section below. For example, the disclosed NOV5 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated NOV5 epitope is from about amino acids 5 to 50. In another embodiment, a NOV5 epitope is from about amino acids 75 to 125. In additional embodiments, NOV5 epitopes are from about amino acids 175 to 225, from about amino acids 280 to 320, from about amino acids 330 to 25 380, and from about amino acids 390 to 440. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV6

30 A disclosed NOV6 nucleic acid of 1020 nucleotides (also referred to as GMAC073364_A_da1) encoding a novel MITSUGUMIN29-like protein is shown in Table 6A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 2-4 and ending with a TAA codon at nucleotides 818-820. Putative untranslated

regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 6A, and the start and stop codons are in bold letters.

Table 6A. NOV6 Nucleotide Sequence (SEQ ID NO:17)

CATGTCCTCGACCGAGAGCGCCGGCCGCACGGCGGACAAGTCGCCGCGCCAGCAGGTGGACC
GCCTACTCGTGGGGCTGCGCTGGCGGCGGCTGGAGGAGCCGCTGGGCTTCATCAAAGTTCTC
CAGTGGCTCTTTGCTATTTTCGCCTTCGGGTCCTGTGGCTCCTACAGCGGGAGACAGGAGC
AATGGTTCGCTGCAACAACGAAGCCAAGGACGTGAGCTCCATCATCGTTGCATTGGCTATC
CCTTCAGGTTGCACCGGATCCAATATGAGATGCCCTCTGCGATGAAGAGTCCAGCTCCAAG
ACCATGCACCTCATGGGGGACTTCTCTGCACCCGCGAGTTCCTCGTGACCCTTGGCATCTT
TTCTTCTTCTATACCATGGCTGCCCTAGTTATCTACCTGCGCTTCCACAACCTCTACACAG
AGAACAACGCTTCCCCTGGTGGACTTCTGTGTGACTGTCTCCTTCACCTTCTTCTGGCTG
GTAGCTGCAGTGCCTGGGGCAAGGGCCTGACCGATGTCAAGGGGGCCACACGACCATCCAG
CTTGACAGCAGCCATGTCAGTGTGCCATGGAGAGGAAGCAGTGTGAGTGCCTGGGGGCCACGC
CCTCTATGGGCTGGCCAACATCTCCGTGCTCTTGGCTTTATCAACTTCTTCTGTGGGCC
GGGAAGTGTGGTTTGTGTTCAAGGAGACCCCGTGGCATGGACAGGGCCAGGGCCAGGACCA
GGACCAGGACAGGACAGGGCCAGGGTCCCAGCCAGGAGTGCAGCTGAGCAGGGAGCAG
TGGAGAAGCAGTAAGCAGCCCCCACCT

The NOV6 nucleic acid was identified on chromosome 3 and has 727 of 805 bases (90%) identical to a gb:GENBANK-ID:AB004816|acc:AB004816.1 mRNA from *Oryctolagus cuniculus* (*Oryctolagus cuniculus* mRNA for mitsugumin29, complete cds ($E = 2.5e^{-142}$)).

A disclosed NOV6 polypeptide (SEQ ID NO:18) encoded by SEQ ID NO:17 is 272 amino acid residues and is presented using the one-letter code in Table 6B. Signal P, Psort and/or Hydropathy results predict that NOV6 has a signal peptide and is likely to be localized on the plasma membrane with a certainty of 0.6000. In other embodiments, NOV6 may also be localized to the Golgi body with a certainty of 0.4000, the endoplasmic reticulum (membrane) with a certainty of 0.3000, or the nucleus with a certainty of 0.1000. The most likely cleavage site for NOV6 is between positions 57 and 58, SYS-GE.

Table 6B. Encoded NOV6 protein sequence (SEQ ID NO:18)

MSSTESAGRTADKSPRQQVDRLLVGLRWRRLEPLGFIKVLQWLFAIFAFGSCGSYSGETGAMVRCNNEAKD
VSSIIVAFGYPFRLHRIQYEMPLCDEESSSKTMHLMGDFSAPAEFFVTLGIFSFYTMALVIYLRFNLYT
ENKRFLVDFCVTVSFTFFWLVAAGWKGKLTGVKATRPSSLTAAAMSVCHGEEAVCSAGATPSMGLANISV
LFGFINFFLWAGNCWFVKETFPWHGQGGQDQDQDQDQGGPSQESAAEQGAVEKQ

The disclosed NOV6 amino acid sequence has 272 of 805 amino acid residues (90%) identical to, and 727 of 805 amino acid residues (90%) similar to, the 3489 amino acid residue gb:GENBANK-ID:AB004816|acc:AB004816.1 protein from *Oryctolagus cuniculus* (*Oryctolagus cuniculus* mRNA for mitsugumin29, complete cds) ($E = 2.5e^{-142}$).

Based on the semi quantitative PCR, NOV6 is specially expressed in: Skeletal muscle, Heart, Kidney, Adrenal gland and one of the Lung cancer cell lines (Lung cancer NCI-H522) at a measurably higher level than the following tissues: Endothelial cells, Pancreas, Thyroid,

Salivary gland, Pituitary gland, Brain (fetal), Brain (whole), Brain (amygdala), Brain (cerebellum), Brain (hippocampus), Brain (thalamus), Cerebral Cortex, Spinal cord, Bone marrow, Thymus, Spleen, Lymph node, Colorectal, Stomach, Small intestine, Bladder, Trachea, Kidney (fetal), Liver, Liver (fetal), Lung, Lung (fetal), Mammary gland, Ovary, Uterus, Placenta, Prostate, Testis, Melanoma, Adipose and cancer cell lines including Breast cancer, CNS cancer, Colon cancer, Gastric cancer, Lung cancer (except Lung cancer NCI-H522), Ovarian cancer, Pancreatic cancer, and Renal cancer.

In addition, NOV6 is predicted to be expressed in skeletal muscle because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AB004816|acc:AB004816.1) a closely related mitsugumin29 homolog in *Oryctolagus cuniculus*.

NOV6 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 6C.

Table 6C. BLAST results for NOV6					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 3077703 dbj BAA25784.1 (AB004816)	mitsugumin29 [Oryctolagus cuniculus]	264	252/272 (92%)	256/272 (93%)	e-136
gi 6678874 ref NP_032622.1 (NM_008596)	mitsugumin 29 [Mus musculus]	264	246/272 (90%)	258/272 (94%)	e-133
gi 12836843 dbj BAB23831.1 (AK005132)	putative [Mus musculus]	285	118/251 (47%)	158/251 (62%)	7e-59
gi 1351168 sp P20488 SYPH_BOVIN	SYNAPTOPHYSIN (MAJOR SYNAPTIC VESICLE PROTEIN P38)	307	109/221 (49%)	145/221 (65%)	3e-58
gi 2134413 pir I50720	synaptophysin IIa - chicken	268	109/217 (50%)	142/217 (65%)	4e-57

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 6D.

Table 6D Clustal W Sequence Alignment

- 1) NOV6 (SEQ ID NO:18)
- 2) gi|3077703|dbj|BAA25784.1| (AB004816) mitsugumin29 [Oryctolagus cuniculus] (SEQ ID NO:57)
- 3) gi|6678874|ref|NP_032622.1| (NM_008596) mitsugumin 29 [Mus musculus] (SEQ ID NO:58)
- 4) gi|12836843|dbj|BAB23831.1| (AK005132) putative [Mus musculus] (SEQ ID NO:59)
- 5) gi|1351168|sp|P20488|SYPH_BOVIN SYNAPTOPHYSIN (MAJOR SYNAPTIC VESICLE PROTEIN P38) (SEQ ID NO:60)
- 6) gi|2134413|pir||I50720 synaptophysin IIa - chicken (SEQ ID NO:61)

```

      10      20      30      40      50
NOV6      ....|....|....|....|....|....|....|....|....|....|
gi|3077703| MSSTESAGRTADKSPRQOVDRLLVGLRWRRLEEPLGFIKVLQWLFAIFAF
gi|6678874| MSSTESPSRAADKSPRQOVDRLLVGLRWRRLEEPLGFIKVLQWLFAIFAF
MSSTESPGRTSDKSPRQOVDRLLVGLRWRRLEEPLGFIKVLQWLFAIFAF

```

		gi 12836843	-----MDPVSVQVASAGTFRALKKEPLAFLRALELLFAMFAF
		gi 1351168	-----MDVVNQLVANGQFFVVKKEPLGFVKVLQWVFAIFAF
		gi 2134413	-----MCMVILFAPLFAIFAF
5			60 70 80 90 100
	NOV6	
	gi 3077703		GS CGSYSGETGAMVRCNNEAKDVSSIIIVAFGYPPRLHRIQYEMPLCDEES
	gi 6678874		GS CGSYSGETGAMVRCNNEAKDVSSIIIVLFGYPPRLHRIEYEMPLCDDDS
10	gi 12836843		GS CGSYSGETGALVLCNNEAKDVSSIIIVLFGYPPRLYOQYEMPLCDQDS
	gi 1351168		ATCGGYSGGRLRLSVDCVNKTESNLSIDTAFAYPRLQOVTFEVPTCEGK-
	gi 2134413		ATCGSYSGELQLSVDCANKTKSDLNIEVEFEYPRLHEVYFEAPTCEG--
			ATCGGYSGGRLRLSVDCANKSESDLNIDTAFAYPRLHQVNFDAPTCEGK-
15			110 120 130 140 150
	NOV6	
	gi 3077703		SSKTMHLMGDFSAPAEFFVTILGIESFFYTMAALVITYLRHNLTYTENKRF
	gi 6678874		SSKTMHLMGDFSAPAEFFVTILGIESFFYTMAALVVITYLRHKLYTENKRF
	gi 12836843		TSKTMHLMGDFSAPAEFFVTILGIESFFYTMAALVITYLRHKLYTENKRF
20	gi 1351168		EQOKLALVGDSSSAEFFVTIVAFALYSLAATVVYIFLQNKYRENNRGP
	gi 2134413		DPKKTFLVGNYSSEAEFFVTIVAFALYSGALATYIFLQNKYRENNRGP
			RRETLSLIGDFSSEAEFFVTIVAFALYSLAATVVYIFLQNKYRENNRGP
25			160 170 180 190 200
	NOV6	
	gi 3077703		LVDFCVTVSFTFFWLVAATAWAGKGLTDVKGATRESSLTAAMSVCHGEEAV
	gi 6678874		LVDFCVTVSFTFFWLVAATAWAGKGLTDVKGATRESSLTAAMSVCHGEEAV
	gi 12836843		LVDFCVTVSFTFFWLVAATAWAGKGLTDVKGATRESSLTAAMSVCHGEEAV
	gi 1351168		LIDFIVTVVFSEFLWLVGSSAWAKGLSDVKVATDEKEVLLLSACKQPSNK
30	gi 2134413		MLDELATAVEFAEMWLVSSEAWAKGLSDVKMATDEENLIKGMHVCHOPGNT
			LIDFIVTVVFSEFLWLVGSSAWAKGLSDVKIATDEDEVLLLSACKQPSNK
35			210 220 230 240 250
	NOV6	
	gi 3077703		CSAGATFSMGLANISVLFGLINFLWAGNCWFVEKETPWHGQGGQDQDQ
	gi 6678874		CSAGATFSMGLANISVLFGLINFLWAGNCWFVEKETPWHGQGGQ-----
	gi 12836843		CSAGATFSMGLANISVLFGLINFLWAGNCWFVEKETPWHGQGGQ-----
	gi 1351168		CMAVHSPVMSSINTSVVFCFLNFIWAGNIWFVEKETGWHSSGORYLSDP
	gi 2134413		CKELRDEVTSGINTSVVFCFLNLVLWVGNLWFVEKETGWAAPFLRAPPGA
40			CLPVRSPVMSSINTSVVFCFLNFIWAGNIWFVEKETGWHSSGQRHAADT
45			260 270 280 290 300
	NOV6	
	gi 3077703		DQDQ-----GQGP-----SQESAAEQG-----
	gi 6678874		--DQ-----GQGP-----SQESAAEQG-----
	gi 12836843		--DQ-----GQGP-----SQESAAEQG-----
	gi 1351168		MEKH-----SSSYNOG--RYN-QESYGSQGGYS-----QQAN-----L
	gi 2134413		PEKQAPAGDAYQAGYGOQPGGYGPODSYGFQGGYQPDYQGPASSGGGGY
			MEKO-----SSGYNOG--GYN-QDSYGPAGGYN-----QPGS-----Y
50			310 320
	NOV6	
	gi 3077703		-----AVEKQ-----
	gi 6678874		-----AVEKQ-----
	gi 12836843		-----AVEKQ-----
55	gi 1351168		GPTSDEFGQQP----SGPTSFNNOI
	gi 2134413		GPQG-DYGQQGYGPQGAPTSFSNOM
			GQVG-DYGQPGSYGQSGPTSFANOI

Table 6E lists the domain description from DOMAIN analysis results against NOV6.

This indicates that the NOV6 sequence has properties similar to those of other proteins known to contain this domain.

Table 6E. Domain Analysis of NOV6

gnl|Pfam|pfam01284, Synaptophysin, Synaptophysin / synaptoporin. (SEQ ID NO:103)
 CD-Length = 298 residues, 70.8% aligned
 Score = 244 bits (622), Expect = 6e-66

5	NOV 5:	29	RRLEEPLGFIKVLQWLFAIFAFGSCGSYSGETGAMVRCNNEAKDVSSIIVAFGYPFRLHR	88
	Sbjct:	3	MFVIFAPLGFVKVLQWVFAIFAFATCGGYSSELQLSVDCANKTESDLNIDIAFAYPFLHE	62
10	NOV 5:	89	IQYEMPLCDEESSKTMHLMGDFSAPAEFFVTTLGIFSFFYTMAALVIYLRFHNLTYENKR	148
	Sbjct:	63	VTFEAPTC-EGDEKKNIALVGDSSSSAEFFVTVAVFALYSLAALATYIFFQNKYRENNK	121
15	NOV 5:	149	FPLVDFCVTVSFTFFWLVAANAAGKGLTDVKGATRPSSSLTAAMSVCHGEEAVCSAGATPS	208
	Sbjct:	122	GPLIDFIATAVFAFLWLVGSSAWAKGLSDVKMATDPEEIIKGMHACHQPGNKCKELHDPV	181
20	NOV 5:	209	MGLANISVLFQFINFFLWAGNCWFVFKETPDH	240
	Sbjct:	182	MSGLNTSVVFGFLNFILWAGNIWVFKETGWA	213

In skeletal muscle, excitation-contraction (E-C) coupling requires the conversion of the depolarization signal of the invaginated surface membrane, namely the transverse (T-) tubule, to Ca²⁺ release from the sarcoplasmic reticulum (SR) (Takeshima H et al., Biochem J 1998 Apr 1;331 (Pt 1):317-22 / PMID: 9512495, UI: 98180964). Signal transduction occurs at the junctional complex between the T-tubule and SR, designated as the triad junction, which contains two components essential for E-C coupling, namely the dihydropyridine receptor as the T-tubular voltage sensor and the ryanodine receptor as the SR Ca²⁺-release channel. However, functional expression of the two receptors seemed to constitute neither the signal-transduction system nor the junction between the surface and intracellular membranes in cultured cells, suggesting that some as-yet-unidentified molecules participate in both the machinery. In addition, the molecular basis of the formation of the triad junction is totally unknown. It is therefore important to examine the components localized to the triad junction. Takeshima et al. report the identification using monoclonal antibody and primary structure by cDNA cloning of mitsugumin29, a novel transmembrane protein from the triad junction in skeletal muscle. This protein is homologous in amino acid sequence and shares characteristic structural features with the members of the synaptophysin family. The subcellular distribution and protein structure suggest that mitsugumin29 is involved in communication between the T-tubular and junctional SR membranes.

Physiological roles of the members of the synaptophysin family, carrying four transmembrane segments and being basically distributed on intracellular membranes including synaptic vesicles, have not been established yet (Nishi M et al., J Cell Biol 1999 Dec

27;147(7):1473-80 / PMID: 10613905, UI: 20082885). Recently, mitsugumin29 (MG29) was identified as a novel member of the synaptophysin family from skeletal muscle. MG29 is expressed in the junctional membrane complex between the cell surface transverse (T) tubule and the sarcoplasmic reticulum (SR), called the triad junction, where the depolarization signal is converted to $\text{Ca}(2+)$ release from the SR. In this study, Nishi et al. examined biological functions of MG29 by generating knockout mice. The MG29-deficient mice exhibited normal health and reproduction but were slightly reduced in body weight. Ultrastructural abnormalities of the membranes around the triad junction were detected in skeletal muscle from the mutant mice, i.e., swollen T tubules, irregular SR structures, and partial misformation of triad junctions. In the mutant muscle, apparently normal tetanus tension was observed, whereas twitch tension was significantly reduced. Moreover, the mutant muscle showed faster decrease of twitch tension under $\text{Ca}(2+)$ -free conditions. The morphological and functional abnormalities of the mutant muscle seem to be related to each other and indicate that MG29 is essential for both refinement of the membrane structures and effective excitation-contraction coupling in the skeletal muscle triad junction. These results further imply a role of MG29 as a synaptophysin family member in the accurate formation of junctional complexes between the cell surface and intracellular membranes.

The temporal appearance and subcellular distribution of mitsugumin29 (MG29), a 29-kDa transmembrane protein isolated from the triad junction in skeletal muscle, were examined by immunohistochemistry during the development of rabbit skeletal muscle (Komazaki S et al., Dev Dyn 1999 Jun;215(2):87-95 / PMID: 10373013, UI: 99300228). MG29 appeared in the sarcoplasmic reticulum (SR) in muscle cells at fetal day 15 before the onset of transverse tubule (T tubule) formation. In muscle cells at fetal day 27, in which T tubule and triad formation is ongoing, both SR and triad were labeled for MG29. In muscle cells at newborn 1 day, the labeling of the SR had become weak and the triads were well developed and clearly labeled for MG29. Specific and clear labeling for MG29 was restricted to the triads in adult skeletal muscle cells. When MG29 was expressed in amphibian embryonic cells by injection of the cRNA, a large quantity of tubular smooth-surfaced endoplasmic reticulum (sER) was formed in the cytoplasm. The tubular sER was 20-40 nm in diameter and appeared straight or reticular in shape. The tubular sER was formed by the fusion of coated vesicles [budded off from the rough-surfaced endoplasmic reticulum (rER)] and vacuoles of rER origin. The present results suggest that MG29 may play important roles both in the formation of the SR and the construction of the triads during the early development of skeletal muscle cells.

Recently mitsugumin29 unique to the triad junction in skeletal muscle was identified as a novel member of the synaptophysin family; the members of this family have four transmembrane segments and are distributed on intracellular vesicles. In this study, Shimuta et al. FEBS Lett 1998 Jul 17;431(2):263-7 / PMID: 9708916, UI: 98372647, isolated and analyzed mouse mitsugumin29 cDNA and genomic DNA containing the gene. The mitsugumin29 gene mapped to the mouse chromosome 3 F3-H2 is closely related to the synaptophysin gene in exon-intron organization, which indicates their intimate relationship in molecular evolution. RNA blot hybridization and immunoblot analysis revealed that mitsugumin29 is expressed abundantly in skeletal muscle and at lower levels in the kidney. Immunofluorescence microscopy demonstrated that mitsugumin29 exists specifically in cytoplasmic regions of the proximal and distal tubule cells in the kidney. The results obtained may suggest that mitsugumin29 is involved in the formation of specialized endoplasmic reticulum systems in skeletal muscle and renal tubule cells.

The disclosed NOV6 nucleic acid of the invention encoding a MITSUGUMIN29 -like protein includes the nucleic acid whose sequence is provided in Table 6A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 6A while still encoding a protein that maintains its MITSUGUMIN29 -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 10 percent of the bases may be so changed.

The disclosed NOV6 protein of the invention includes the MITSUGUMIN29 -like protein whose sequence is provided in Table 6B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 6B while still encoding a protein that maintains its MITSUGUMIN29-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 10 percent of the residues may be so changed.

The NOV6 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in eye/lens disorders including but not limited to muscular dystrophy, Lesch-Nyhan syndrome, myasthenia gravis, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, and other diseases, disorders and conditions of the like. Also since the invention is highly expressed in one of the lung cancer cell lines (Lung cancer NCI-H522), it may be useful in diagnosis and treatment of this cancer. The NOV6 nucleic acid, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV6 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV6 protein have multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, contemplated NOV6 epitope is from about amino acids 10 to 40. In other embodiments, NOV6 epitope is from about amino acids 60 to 70, from about amino acids 90 to 130, from about amino acids 145 to 155, from about amino acid 170 to 180, and from about amino acids 220 to 270. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV7

A disclosed NOV7 nucleic acid of 1020 nucleotides (also referred to as 106973211_EXT) encoding a novel Wnt-15-like protein is shown in Table 7A. An open reading frame was identified beginning with an CTG initiation codon at nucleotides 2-4 and ending with a TAG codon at nucleotides 995-997. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 7A, and the start and stop codons are in bold letters. Since the starting codon is not a

traditional initiation codon, NOV7 could represent a partial reading frame, and could further extend in the 5' direction.

Table 7A. NOV7 Nucleotide Sequence (SEQ ID NO:19)

```
CCTGACCGGGCGGGAAGTCCTGACGCCCTTCCCAGGATTGGGCACTGCGGCAGCCCCGGCACAGGGCGGG
GCCCACCTGAAGCAGTGTGACCTGCTGAAGCTGTCCCGGGCGGCAGAAGCAGCTCTGCCGGAGGGAGCCCCG
GCCTGGCTGAGACCTGAGGGATGCTGCGCACCTCGGCCCTGCTTGAGTGCCAGTTTCAGTTCCGGCATGA
GCGCTGGAACTGTAGCCTGGAGGGCAGGATGGGCTGCTCAAGAGAGGCTTCAAAGAGACAGCTTTCCTG
TACGCGGTGTCTCTGCGGCCCTCACCACACCTGGCCCGGGCCTGCAGCGCTGGGCGCATGGAGCGCT
GCACCTGTGATGACTCTCGGGGCTGGAGAGCCCGCAGGCTGGCAGTGGGGCGTGTGCGGTGACAACT
CAAGTACAGCACCAAGTTTCTGAGCAACTTCTGGGGTCCAAGAGAGGAAACAAGGACCTGCGGGCACGG
GCAGACGCCCCACAATACCCACGTGGGCATCAAGGCTGTGAAGAGTGGCCTCAGGACCACGTGTAAGTGCC
ATGGCGTATCAGGCTCCTGTGCGCTGCGCACCTGTGGAAGCAGCTCTCCCGTTCCGTGAGACGGGCCA
GGTGCTGAAACTGCGCTATGACTCGGCTGTCAAGTGTCCAGTGCCACCAATGAGGCCTTGGGCGCGCTA
GAGCTGTGGGCCCCCTGCCAGGCAGGGCAGCCTCACAAAGGCCTGGCCCCAAGGCTGGGGACCTGGTGT
ACATGGAGGACTCACCCAGCTTCTGCCGGCCAGCAAGTACTCACTGGCACAGCAGGTAGGGTGTGCTC
CCGGGAGGCCAGTGCAGCAGCCTGTGCTGCGGGCGGGGCTATGACACCCAGAGCCGCTGTTGGCCTTC
TCTGCCACTGCCAGGTGCAGTGGTGTGCTACGTGGAGTGCCAGCAATGTGTGAGGAGGAGCTTGTGT
ACACCTGCAAGCACTAGGCCTACTGCCAGCAAGCCAGTC
```

The disclosed NOV7 nucleic acid sequence, located on chromosome 17, has 688 of 1009 bases (68%) identical to a gb:GENBANK-ID:AF031168|acc:AF031168.1 mRNA from *Gallus gallus* (*Gallus gallus* Wnt-14 protein (Wnt-14) mRNA, complete cds) ($E = 3.0e^{-76}$).

A disclosed NOV7 polypeptide (SEQ ID NO:20) encoded by SEQ ID NO:19 is 331 amino acid residues and is presented using the one-letter amino acid code in Table 7B. Signal P, Psort and/or Hydropathy results predict that NOV7 contains no signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.4500. In other embodiments, NOV7 is also likely to be localized to the microbody (peroxisome) with a certainty of 0.3000, the mitochondrial matrix space with a certainty of 0.1000, or to the lysosome (lumen) with a certainty of 0.1000.

Table 7B. Encoded NOV7 protein sequence (SEQ ID NO:20).

```
LTGREVLTPFPLGLTAAAPAQGGAHLKQCDLLKLSRRQKQLCRREPGLAETLRDAAHLGLLECCQFQFRHERWNCS
LEGRMGLLKRGFKETAFLYAVSSAALHTHTLARACSAGRMRCTCDDSPGLSRQAWQWGVCGDNLKYSTKFLSNF
LGSKRGNKDLRARADAHNTHVGIKAVKSLRRTCKCHGVSGSCAVRTCWKQLSPFRETGQVLKRLRYDSAVKVSSA
TNEALGRLELWAPARQGSLLTKGLAPRSGDLVYMEDSPSFCRPSKYSPTAGRVCSREASCSSLCCGRGYDTQSRL
VAFSCHCQVQWCCYVECCQCVQEELVYTCCKH
```

The disclosed NOV7 amino acid sequence has 205 of 330 amino acid residues (62%) identical to, and 252 of 330 amino acid residues (76%) similar to, the 354 amino acid residue ptnr:SWISSPROT-ACC:O42280 protein from *Gallus gallus* (Chicken) (WNT-14 Protein Precursor) ($E = 1.3e^{-114}$).

The tissue expression of NOV7 is predicted to be expressed in brain because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AF031168|acc:AF031168.1) a closely related *Gallus gallus* Wnt-14 protein (Wnt-14) mRNA, complete cds homolog.

NOV7 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 7C.

Table 7C. BLAST results for NOV7					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 16303264 dbj BAB70499.1 (AB063483)	WNT14B [Homo sapiens]	357	330/331 (99%)	330/331 (99%)	e-175
gi 3915306 sp O42280 WN14 CHICK	WNT-14 PROTEIN PRECURSOR	354	204/332 (61%)	253/332 (75%)	e-109
gi 15082261 ref NP_003386.1 (NM_003395)	wingless-type MMTV integration site family, member 14 [Homo sapiens]	365	209/335 (62%)	255/335 (75%)	e-108
gi 139748 sp P10108 WNT1_XENLA	WNT-1 PROTEIN PRECURSOR (XWNT-1) (XINT-1)	371	120/313 (38%)	175/313 (55%)	5e-58
gi 3024851 sp O14905 WN15 HUMAN	WNT-15 PROTEIN	120	120/120 (100%)	120/120 (100%)	2e-56

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 7D.

Table 7D Information for the ClustalW proteins

- 1) NOV7 (SEQ ID NO:20)
- 2) gi|16303264|dbj|BAB70499.1| (AB063483) WNT14B [Homo sapiens] (SEQ ID NO:62)
- 3) gi|3915306|sp|O42280|WN14_CHICK WNT-14 PROTEIN PRECURSOR (SEQ ID NO:63)
- 4) gi|15082261|ref|NP_003386.1| (NM_003395) wingless-type MMTV integration site family, member 14 [Homo sapiens] (SEQ ID NO:64)
- 5) gi|139748|sp|P10108|WNT1_XENLA WNT-1 PROTEIN PRECURSOR (XWNT-1) (XINT-1) (SEQ ID NO:65)
- 6) gi|3024851|sp|O14905|WN15_HUMAN WNT-15 PROTEIN (SEQ ID NO:66)

		10	20	30	40	50
NOV7					
gi 16303264	-----MRPPPALALAGLCLLALPAAASFYGLTGRE-----VLTFFPFGGL					
gi 3915306	-----MALLRALLG--LLACTPRPSAAYFGLTGNE-----ALTILE-L					
gi 15082261	MLDGSPLARWLAAAFGLTLLLAALRPSAAYFGLTGSE-----PLTILF-L					
gi 139748	-----MRILTFLGLKTLWVLAFFSSLSNTIAVNNSGKWWGIVNVASAG					
gi 3024851	-----					
		60	70	80	90	100
NOV7					
gi 16303264	GTAAAPAQGGAHLLKQCDLLKLSRRQKQLCRRREPGLAETLRDAAHLLGLLEC					
gi 3915306	GTAAAPAQGGAHLLKQCDLLKLSRRQKQLCRRREPGLAETLRDAAHLLGLLEC					
gi 15082261	TSEMEBAAVKAAHVKVCDRLKLEKKORRMCRDPGGAETLMEALISMSALEC					
gi 139748	TLEP-EAQAQAHVKACDRLLKLERKORRMCRDPGVAETLVEAVMSALEC					
gi 3024851	NVLPGSDARPVPLVLDPSLQILSRQKRLTRONPGILQSLTRGLHSAIREC					
					
		110	120	130	140	150
NOV7					
	QEQFRHERWNCSELEG--RMG--LLKRGFKETAFLYAVSSAALTHTLARAC					

5	gi 16303264 gi 3915306 gi 15082261 gi 139748 gi 3024851	QFQFRHERWNCSELEG--RTG--LLKRGFKETAFLYAVSSAALHTLARAC QYQFRFERWNCLEGRYRAS--LLKRGFKETAFLYAISSACLTHAMAKAC QFQFRFERWNCLEGRYRAS--LLKRGFKETAFLYAISSACLTHAMAKAC KWHFRNRNRWNCPTGTGNQVFGKIINRGCRETAFFVSAITSAVTHSVARSC -----
10	NOV7 gi 16303264 gi 3915306 gi 15082261 gi 139748 gi 3024851	<div>160 170 180 190 200</div> <div>..... </div> SAGRMERCTCDDSEGLERQAWQWGVCGDNLKYSTKFLSNFLGSKRGNKD SAGRMERCTCDDSEGLERQAWQWGVCGDNLKYSTKFLSNFLGSKRGNKD SAGRMERCTCDEAFDLENREAWQWGC CGDNLKYSNKEVKEFLG-RKPNKD SAGRMERCTCDEAFDLENREAWQWGC CGDNLKYSSKEVKEFLG-RSSKD SEGSIIESCSCDYRRRGPGGPDWHWGCSDNLEFGRFIGREIVDSSERGRD -----
15	NOV7 gi 16303264 gi 3915306 gi 15082261 gi 139748 gi 3024851	<div>210 220 230 240 250</div> <div>..... </div> LRARADAHNTHVGITKAVKSGLRITCKCHGVSGSCAVRTCWQLSPFRETG LRARADAHNTHVGITKAVKSGLRITCKCHGVSGSCAVRTCWQLSPFRETG LRARVDHNNLVGMKVIKAGVETTCCKCHGVSGSCAVRTCWQLSPFHEIG LRARVDHNNLVGMKVIKAGVETTCCKCHGVSGSCAVRTCWQLAPHEVIG LKYLVDHNNLVGMKVIKAGVETTCCKCHGVSGSCAVRTCWQLAPHEVIG -----SGSCAVRTCWQLSPFRETG
25	NOV7 gi 16303264 gi 3915306 gi 15082261 gi 139748 gi 3024851	<div>260 270 280 290 300</div> <div>..... </div> QVLKLRYSADVKSATNEALGRLELWAPAP---QGLTKGLAPRSGDLV QVLKLRYSADVKSATNEALGRLELWAPAP---QGLTKGLAPRSGDLV KOLKOKYETSLKVGSTTNEATGE-GDISPPK--KSI PGHSDQIPRTDLV KHLKHKYETSLKVGSTTNEATGEAGAISSPPGRASAGGSDPLPRIPELV DALKDRFDGASKVTYSNNGSNRWGSRSDPPH--LEPENPTHALPSSQDLV QVLKLRYSADVKSATNEALGRLELWAPAP---QGLTKGLAPRSGDLV -----
35	NOV7 gi 16303264 gi 3915306 gi 15082261 gi 139748 gi 3024851	<div>310 320 330 340 350</div> <div>..... </div> YMEDSPSFQCRPSKYS--PGTAGRVCSRE---ASCSSLCCGRGYDTQSRV YMEDSPSFQCRPSKYS--PGTAGRVCSRE---ASCSSLCCGRGYDTQSRV YIDDSPSFCLMSRYS--PGTSGRKYKID---KNCDSSICCGRGHTQSRV HIDDSPSFCLAGRFSS--PGTAGRRCHRE---KNCESSICCGRGHTQSRV YFEKSPNFQCRPSKYS--PGTTGRICNSTSLGLDCELLCCGRGYRSLAEK YMEDSPSFQCRPSKYS--PGTAGRVCSRE---ASCSSLCCGRGYDTQSRV -----
45	NOV7 gi 16303264 gi 3915306 gi 15082261 gi 139748 gi 3024851	<div>360 370 380</div> <div>..... </div> VAFSCHCQVQWCCYVECCQCVQBELVYTCKH VAFSCHCQVQWCCYVECCQCVQBELVYTCKH VTRPCQCVQWCCYVECCQCVQBELVYTCKD VTRPCQCVQWCCYVECCQCVQBELVYTCKG VTRCHCTFNWCCYVECCQCVQBELVYTCKL VAFSCHCQV-----

Tables 7E and 7F list the domain descriptions from DOMAIN analysis results against NOV7. This indicates that the NOV7 sequence has properties similar to those of other proteins known to contain this domain.

Table 7E. Domain Analysis of NOV7

gnl|Pfam|pfam00110, wnt, wnt family. (SEQ ID NO:104)
CD-Length = 313 residues, 97.8% aligned
Score = 268 bits (684), Expect = 5e-73

	NOV 6:	34	LSRRQKQLCRREPGLAETLRDAHLGLLECGQFQFRHERWNCSELEGRMGL-----LKRGFK	88
			+ + + ++ + + + + ++ +	
	Sbjct:	8	LSPRQRQLCRRNPDMASVSEGAQLAIQECQHFRGRRWNCSTLDRLRVVFGKVLKKGTR	67
5	NOV 6:	89	ETAFLYAVSSAALHTTLARACSAGRMERCTCDDSPG-LESQAWQWGVCGDNLKYSTKFL	147
			+ + + + + + + ++ +	
	Sbjct:	68	ETAFLVYAISSAGVAHAVTRACSEGELESCGCDYKKGPGGPQGSWQWGGCSDNVEFGIRFS	127
	NOV 6:	148	SNFLGSKRGNKDLRARADAHNTHVGIAKAVKSGRLTTCKCHGVSGSCAVRTCWKQLSPFRE	207
10			+ ++ + + + + + + +	
	Sbjct:	128	REFVDAREERERDARSLMNLHNNEAGRKAVKSHMRRECKCHGVSGSCSMKTCWLSLPDFRA	187
	NOV 6:	208	TGQVLKRLYDSAVKVSSATNEALGRLELWAPARQGSITKGLAPRSGDLVYMEDSPSFCR-	266
			+ + + + + + + +	
15	Sbjct:	188	VGDALKDKYDGAIRV---EPNKRGMGQGSAPRLVAKNPRFKPPTRSDLVYLEDSPDYCER	244
	NOV 6:	267	-PSKYSPGTAGRVCSREA----SCSSLCCGRGYDTQSRLVAFSCHCQVQWCCYVECCQCV	321
			+ + + + + + +	
	Sbjct:	245	DRSTGSLGTQGRVCNKTSGKLDGCELLCCGRGYNTQQVERTEKCNCKFWCCYVKCEECQ	304
20				
	NOV 6:	322	QEELVYTCK 330	
			+	
	Sbjct:	305	EVVEVHTCK 313	

Table 7F. Domain Analysis of NOV7

gnl|Smart|smart00097, WNT1, found in Wnt-1 (SEQ ID NO:105)
 CD-Length = 304 residues, 98.7% aligned
 Score = 248 bits (632), Expect = 5e-67

25	NOV 6:	34	LSRRQKQLCRREPGLAETLRDAHLGLLECGQFQFRHERWNCSELEGRMGL-----LKRGFKE	89
			+ + + ++ + + + + ++ +	
	Sbjct:	5	LSRRQRQLCRANPDVMASVAEGAQEGIEECQHFRFRRWNCSTAGLASIFGKVLROGTRE	64
30	NOV 6:	90	TAFLYAVSSAALHTTLARACSAGRMERCTCDDSPGLESRQAWQWGVCGDNLKYSTKFLSN	149
			+ + + + + ++ + + ++ +	
	Sbjct:	65	TAFVYAISSAGVAHAVTRACSQEELDSCGCDYSKRGSGGRGWEGGCSNIDFGIGFSRE	124
35	NOV 6:	150	FLGSK-RGNKDLRARADAHNTHVGIAKAVKSGRLTTCKCHGVSGSCAVRTCWKQLSPFRET	208
			+ ++ + + + + + +	
	Sbjct:	125	FVDARERRGSDARALMNLHNNEAGRLAVKTKMKRECKCHGVSGSCSVKTCWLQLPFIREI	184
	NOV 6:	209	GQVLKRLYDSAVKVSSATNEALGRLELWAPARQGSITKGLAPRSGDLVYMEDSPSFC--R	266
			+ + + + + + + +	
40	Sbjct:	185	GDYLKEKYDGASEV-VLDKRGTGRGLVPANRDFK-----PPTNTDLVYLESSPDFCEKN	236
	NOV 6:	267	PSKYSPGTAGRVCSREA----SCSSLCCGRGYDTQSRLVAFSCHCQVQWCCYVECCQCVQ	322
			+ + + + + + +	
	Sbjct:	237	PKTGSLSLTQGRVCNKTSGKLDGCDLLCCGRGYNTEHVEVVERCNCKFWCCYVKCKQCRE	296
45	NOV 6:	323	EELVYTCK	330
			+	
	Sbjct:	297	RVEKHTCK	304

50 Wnt proteins constitute a large family of molecules involved in cell proliferation, cell differentiation and embryonic patterning. They are known to interact with the Frizzled family of receptors to activate two main intracellular signaling pathways regulating intracellular calcium levels and gene transcription. Early studies on Wnts implicated them in cell proliferation and tumorigenesis, which have been borne out by recent work using transgenic

and null mutant mice. Wnts are involved in processes involved in mammary gland development and cancer. Recent studies have demonstrated that these molecules are critical to organogenesis of several systems, such as the kidney and brain. Wnts regulate the early development, i.e. neural induction, and their role persists in later stages of development as well as in the mature organ. An example of this is seen in the brain, where the loss of certain Wnts leads to the absence of critical regions of the brain, e.g. the hippocampus, involved in learning and memory, or the cerebellum, involved in motor function. Wnts have also been implicated in the genesis of degenerative diseases such as Alzheimer's disease. The protein encoded by the novel gene described herein may therefore play a role in cellular proliferation, differentiation, dysregulation, organogenesis and disease processes such as cancer, developmental defects etc.

A partial sequence corresponding to this novel protein, with homology to the chicken Wnt-14, has been deposited in GenBank with the nomenclature Wnt-15.

Alzheimer's disease (AD) is a neurodegenerative disease with progressive dementia accompanied by three main structural changes in the brain: diffuse loss of neurons; intracellular protein deposits termed neurofibrillary tangles (NFT) and extracellular protein deposits termed amyloid or senile plaques, surrounded by dystrophic neurites. Two major hypotheses have been proposed in order to explain the molecular hallmarks of the disease: The 'amyloid cascade' hypothesis and the 'neuronal cytoskeletal degeneration' hypothesis. While the former is supported by genetic studies of the early-onset familial forms of AD (FAD), the latter revolves around the observation in vivo that cytoskeletal changes - including the abnormal phosphorylation state of the microtubule associated protein tau - may precede the deposition of senile plaques. Recent studies have suggested that the trafficking process of membrane associated proteins is modulated by the FAD-linked presenilin (PS) proteins, and that amyloid beta-peptide deposition may be initiated intracellularly, through the secretory pathway. Current hypotheses concerning presenilin function are based upon its cellular localization and its putative interaction as macromolecular complexes with the cell-adhesion/signaling beta-catenin molecule and the glycogen synthase kinase 3beta (GSK-3beta) enzyme. Developmental studies have shown that PS proteins function as components in the Notch signal transduction cascade and that beta-catenin and GSK-3beta are transducers of the Wnt signaling pathway. Both pathways are thought to have an important role in brain development, and they have been connected through Dishevelled (Dvl) protein, a known transducer of the Wnt pathway.

Members of the vertebrate Wnt family have been subdivided into two functional classes according to their biological activities. Some Wnts signal through the canonical Wnt-1/wingless pathway by stabilizing cytoplasmic beta-catenin. By contrast other Wnts stimulate intracellular Ca^{2+} release and activate two kinases, CamKII and PKC, in a G-protein-dependent manner. Moreover, putative Wnt receptors belonging to the Frizzled gene family have been identified that preferentially couple to the two prospective pathways in the absence of ectopic Wnt ligand and that might account for the signaling specificity of the Wnt pathways. As Ca^{2+} release was the first described feature of the noncanonical pathway, and as Ca^{2+} probably plays a key role in the activation of CamKII and PKC, Kuhl M, et al., (*Trends Genet* 2000 Jul;16(7):279-83) have named this Wnt pathway the Wnt/ Ca^{2+} pathway.

Many constituents of Wnt signaling pathways are expressed in the developing and mature nervous systems. Recent work has shown that Wnt signaling controls initial formation of the neural plate and many subsequent patterning decisions in the embryonic nervous system, including formation of the neural crest. Wnt signaling continues to be important at later stages of development. Wnts have been shown to regulate the anatomy of the neuronal cytoskeleton and the differentiation of synapses in the cerebellum. Wnt signaling has been demonstrated to regulate apoptosis and may participate in degenerative processes leading to cell death in the aging brain.

Recent genetic studies have shown that the signalling factor Wnt3a is required for formation of the hippocampus; the developmental consequences of Wnt signalling in the hippocampus are mediated by multiple HMG-box transcription factors, with LEF-1 being required just for formation of the dentate gyrus.

Wnt-1 was first identified as a protooncogene activated by viral insertion in mouse mammary tumors. Transgenic expression of this gene using a mouse mammary tumor virus LTR enhancer causes extensive ductal hyperplasia early in life and mammary adenocarcinomas in approximately 50% of the female transgenic (TG) mice by 6 months of age. Metastasis to the lung and proximal lymph nodes is rare at the time tumors are detected but frequent after the removal of the primary neoplasm. The potent mitogenic effect mediated by Wnt-1 expression does not require estrogen stimulation; tumors form after an increased latency in estrogen receptor alpha-null mice. Several genetic lesions, including inactivation of p53 and over-expression of Fgf-3, collaborate with Wnt-1 in leading to mammary tumors, but loss of Sky and inactivation of one allele of Rb do not affect the rate of tumor formation in Wnt-1 TG mice.

Communication between cells is often mediated by secreted signaling molecules that bind cell surface receptors and modulate the activity of specific intracellular effectors. The Wnt family of secreted glycoproteins is one group of signaling molecules that has been shown to control a variety of developmental processes including cell fate specification, cell proliferation, cell polarity and cell migration. In addition, mis-regulation of Wnt signaling can cause developmental defects and is implicated in the genesis of several human cancers. The importance of Wnt signaling in development and in clinical pathologies is underscored by the large number of primary research papers examining various aspects of Wnt signaling that have been published in the past several years.

Reproductive tract development and function is regulated by circulating steroid hormones. In the mammalian female reproductive tract, estrogenic compounds direct many aspects of cytodifferentiation including uterine gland formation, smooth muscle morphology, and epithelial differentiation. While it is clear that these hormones act through their cognate nuclear receptors, it is less clear what signaling events follow hormonal stimulation that govern cytodifferentiation. Recent advances in molecular embryology and cancer cell biology have identified the Wnt family of secreted signaling molecules. Discussed here are recent advances that point to a definitive role during uterine development and adult function for one member of the Wnt gene family, Wnt-7a. In addition, recent data is reviewed that implicates Wnt-7a deregulation in response to pre-natal exposure to the synthetic estrogenic compound, DES. These advances point to an important role for the Wnt gene family in various reproductive tract pathologies including cancer.

Holoprosencephaly (HPE) is the most common developmental defect of the forebrain in humans. Several distinct human genes for holoprosencephaly have now been identified. They include Sonic hedgehog (SHH), ZIC2, and SIX3. Many additional genes involved in forebrain development are rapidly being cloned and characterized in model vertebrate organisms. These include Patched (Ptc), Smoothened (Smo), cubitus interruptus (ci)/Gli, wingless (wg/Wnt, decapentaplegic (dpp)/BMP, Hedgehog interacting protein (Hip), nodal, Smads, One-eyed pinhead (Oep), and TG-Interacting Factor (TGIF). However, further analysis is needed before their roles in HPE can be established.

Female reproductive hormones control mammary gland morphogenesis. In the absence of the progesterone receptor (PR) from the mammary epithelium, ductal side-branching fails to occur. Briskin C, et al. (*Genes Dev* 2000 Mar 15;14(6):650-4) overcame this defect by ectopic expression of the protooncogene Wnt-1. Transplantation of mammary epithelia from Wnt-4(-)/(-) mice shows that Wnt-4 has an essential role in side-branching early in pregnancy. PR and

Wnt-4 mRNAs colocalize to the luminal compartment of the ductal epithelium. Progesterone induces Wnt-4 in mammary epithelial cells and is required for increased Wnt-4 expression during pregnancy. Thus, Wnt signaling is essential in mediating progesterone function during mammary gland morphogenesis.

5 Synapse formation requires changes in cell morphology and the upregulation and localization of synaptic proteins. In the cerebellum, mossy fibers undergo extensive remodeling as they contact several granule cells and form complex, multisynaptic glomerular rosettes. Hall AC, et al., (*Cell* 2000 Mar 3;100(5):525-35) showed that granule cells secrete factors that induce axon and growth cone remodeling in mossy fibers. This effect is blocked by the
10 WNT antagonist, sFRP-1, and mimicked by WNT-7a, which is expressed by granule cells. WNT-7a also induces synapsin I clustering at remodeled areas of mossy fibers, a preliminary step in synaptogenesis. Wnt-7a mutant mice show a delay in the morphological maturation of glomerular rosettes and in the accumulation of synapsin I. We propose that WNT-7a can function as a synaptogenic factor.

15 Estrogens have important functions in mammary gland development and carcinogenesis. To better define these roles, Bocchinfuso WP, et al., (*Cancer Res* 1999 Apr 15;59(8):1869-76) have used two previously characterized lines of genetically altered mice: estrogen receptor-alpha (ER alpha) knockout (ERKO) mice, which lack the gene encoding ER alpha, and mouse mammary virus tumor (MMTV)-Wnt-1 transgenic mice (Wnt-1 TG), which
20 develop mammary hyperplasia and neoplasia due to ectopic production of the Wnt-1 secretory glycoprotein. Bocchinfuso WP, et al. have crossed these lines to ascertain the effects of ER alpha deficiency on mammary gland development and carcinogenesis in mice expressing the Wnt-1 transgene. Introduction of the Wnt-1 transgene into the ERKO background stimulates proliferation of alveolar-like epithelium, indicating that Wnt-1 protein can promote
25 mitogenesis in the absence of an ER alpha-mediated response. The hyperplastic glandular tissue remains confined to the nipple region, implying that the requirement for ER alpha in ductal expansion is not overcome by ectopic Wnt-1. Tumors were detected in virgin ERKO females expressing the Wnt-1 transgene at an average age (48 weeks) that is twice that seen in virgin Wnt-1 TG mice (24 weeks) competent to produce ER alpha. Prepubertal ovariectomy of
30 Wnt-1 TG mice also extended tumor latency to 42 weeks. However, pregnancy did not appear to accelerate the appearance of tumors in Wnt-1 TG mice, and tumor growth rates were not measurably affected by late ovariectomy. Small hyperplastic mammary glands were observed in Wnt-1 TG males, regardless of ER alpha gene status; the glands were similar in appearance to those found in ERKO/Wnt-1 TG females. Mammary tumors also occurred in Wnt-1 TG

males; latency tended to be longer in the heterozygous ER alpha and ERKO males (86 to 100 weeks) than in wild-type ER alpha mice (ca. 75 weeks). Bocchinfuso WP, et al. concluded that ectopic expression of the Wnt-1 proto-oncogene can induce mammary hyperplasia and tumorigenesis in the absence of ER alpha in female and male mice. The delayed time of tumor appearance may depend on the number of cells at risk of secondary events in the hyperplastic glands, on the carcinogenesis-promoting effects of ER alpha signaling, or on both.

Wnt-1 and Wnt-3a proto-oncogenes have been implicated in the development of midbrain and hindbrain structures. Evidence for such a role has been derived from in situ hybridization studies showing Wnt-1 and -3a expression in developing cranial and spinal cord regions and from studies of mutant mice whose Wnt-1 genes have undergone targeted disruption by homologous recombination. Wnt-1 null mutants exhibit cranial defects but no spinal cord abnormalities, despite expression of the gene in these regions. The absence of spinal cord abnormalities is thought to be due to a functional compensation of the Wnt-1 deficiency by related genes, a problem that has complicated the analysis of null mutants of other developmental genes as well. Augustine K, et al., (*Dev Genet* 1993;14(6):500-20) describe the attenuation of Wnt-1 expression using antisense oligonucleotide inhibition in mouse embryos grown in culture. Augustine K, et al. induced similar mid- and hindbrain abnormalities as those seen in the Wnt-1 null mutant mice. Attenuation of Wnt-1 expression was also associated with cardiomegaly resulting in hemostasis. These findings are consistent with the possibility that a subset of Wnt-1 expressing cells include neural crest cells known to contribute to septation of the truncus arteriosus and to formation of the visceral arches. Antisense knockout of Wnt-3a, a gene structurally related to Wnt-1, targeted the forebrain and midbrain region, which were hypoplastic and failed to expand, and the spinal cord, which exhibited lateral outpocketings at the level of the forelimb buds. Dual antisense knockouts of Wnt-1 and Wnt-3a targeted all brain regions leading to incomplete closure of the cranial neural folds, and an increase in the number and severity of outpocketings along the spinal cord, suggesting that these genes complement one another to produce normal patterning of the spinal cord. The short time required to assess the mutant phenotype (2 days) and the need for limited sequence information of the target gene (20-25 nucleotides) make this antisense oligonucleotide/whole embryo culture system ideal for testing the importance of specific genes and their interactions in murine embryonic development.

Wnt-1 (previously known as int-1) is a proto-oncogene induced by the integration of the mouse mammary tumor virus. It is thought to play a role in intercellular communication and seems to be a signalling molecule important in the development of the central nervous

system (CNS). The sequence of wnt-1 is highly conserved in mammals, fish, and amphibians. Wnt-1 is a member of a large family of related proteins that are all thought to be developmental regulators. These proteins are known as wnt-2 (also known as irp), wnt-3 up to wnt-15. At least four members of this family are present in *Drosophila*. One of them, wingless (wg), is implicated in segmentation polarity. All these proteins share the following features characteristics of secretory proteins, a signal peptide, several potential N-glycosylation sites and 22 conserved cysteines that are probably involved in disulfide bonds. The Wnt proteins seem to adhere to the plasma membrane of the secreting cells and are therefore likely to signal over only few cell diameters.

The disclosed NOV7 nucleic acid of the invention encoding a Wnt-15-like protein includes the nucleic acid whose sequence is provided in Table 7A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 7A while still encoding a protein that maintains its Wnt-15-like activities and physiological functions, or a fragment of such a nucleic acid.

The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 32 percent of the bases may be so changed.

The disclosed NOV7 protein of the invention includes the Wnt-15-like protein whose sequence is provided in Table 7B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 7B while still encoding a protein that maintains its Wnt-15-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 38 percent of the residues may be so changed.

The above defined information for this invention suggests that these Wnt-15-like proteins (NOV7) may function as a member of a "Wnt-15 family". Therefore, the NOV7 nucleic acids and proteins identified here may be useful in potential therapeutic applications

implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The nucleic acids and proteins of NOV7 are useful in Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, cancer, developmental defects, and/or other pathologies and disorders. The novel NOV7 nucleic acid encoding NOV7 protein, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

NOV7 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV7 protein have multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, contemplated NOV7 epitope is from about amino acids 25 to 60. In other embodiments, NOV7 epitope is from about amino acids 65 to 80, from about amino acids 110 to 140, from about amino acids 145 to 180, from about amino acids 190 to 220, from about amino acids 230 to 270, or from about amino acids 280 to 290. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV8

A disclosed NOV8 nucleic acid of 1085 nucleotides (also referred to 88091010_EXT) encoding a novel Wnt-14-like protein is shown in Table 8A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 13-15 and ending with a

TGA codon at nucleotides 1078-1080. In Table 8A, the 5' and 3' untranslated regions are underlined and the start and stop codons are in bold letters.

Table 8A. NOV8 Nucleotide Sequence (SEQ ID NO:21)

TAGTGAGCCGAGATGGCACTACTATATTCAGCTTGGGTGTGGTGTGTGTGCACCTGTAGTCCTAGTTACTT
 TGGACTGACGGGCAGCGAGCCCCTGACCATCCTCCCGCTGACCTGGAGCCAGAGCGGGCTGCCAGGCGC
 ACTACAAGGCCCTGCGACCGGCTGAAGCTGGAGCGGAAGCAGCGGCGCATGTGCCGCCGGGACCCGGGCGTG
 GCAGAGACGCTGGTGGAGGCCGTGAGCATGAGTGGCTCGCTCGAGTGCCAGTTCCAGTTCCGCTTTGAGCGCTG
 GAACTGCACGCTGGAGGGCCGCTACCGGGCCAGCCTGCTCAAGCGAGGTTTCAAGGAGACTGCCTTCTCTCT
 ATGCCATCTCCTCGGCTGGCCTGACGCAACGCTGGCCCAAGGCGTGCAGCGCGGGCCGCTGGAGCGCTGT
 ACCTGCGATGAGGCACCCGACCTGGAGAACCCTGAGGGCTGGAAGTGGGGTGGCTGTAGCGAGGACATCGA
 GTTTGGTGGGATGGTGTCTCGGGAGTTCGCCGACGCCCGGGAGAACCGGCCAGATGCCCGCTCAGCCATGA
 ACCGCCACAACAACGAGGCTGGGCGCCAGGTGATCAAGGCTGGGGTGGAGACCACCTGCAAGTGCCACGGC
 GTGTGAGGCTCATGCACGGTGGCGACCTGCTGGCGGCAGTTGGCGCCTTTCCATGAGGTGGGCAAGCATCT
 GAAGCACAAGTATGAGTCGGCACTCAAGTGCGGCGAGCACCACCAATGAAGTGCCTGGCGGAGGCGAGTGCCA
 TCTCCCCACACGGGGCGTGCCTCGGGGCGAGTGGCGAGCGACCGCTGCCCGCACTCCAGAGCTGGTG
 CACCTGGATGACTCGCCTAGCTTCTGCTGGCTGGCGGCTTCTCCCCGGGCACCGTGGCCGTAGGTGCCA
 CCGTGAGAAGAACTGCGAGAGCATCTGCTGTGGCGCGGCCATAACACACAGAGCCGGGTGGTGACAAGGC
 CCGTGCAGTGCCAGGTGCGTTGGTGTCTATGTGGAGTGCAGGAGTGCACGACGCTGAGGAGGTCTAC
 ACCTGCAAGGGCTGAGTTC

5 The disclosed NOV8 nucleic acid sequence, localized to chromosome 1, has 560 of 725 bases (77%) identical to a gb:GENBANK-ID:AF031168|acc:AF031168.1 mRNA from *Gallus gallus* (*Gallus gallus* Wnt-14 protein (Wnt-14) mRNA, complete cds (E = 5.2e⁻¹¹⁵).

A disclosed NOV8 polypeptide (SEQ ID NO:22) encoded by SEQ ID NO:21 is 355 amino acid residues and is presented using the one-letter amino acid code in Table 8B. Signal
 10 P, Psort and/or Hydropathy results predict that NOV8 has a signal peptide and is likely to be localized extracellularly with a certainty of 0.3700. In other embodiments, NOV8 is also likely to be localized to the endoplasmic reticulum (membrane) with a certainty of 0.1000, to the endoplasmic reticulum (lumen) with a certainty of 0.1000, or the lysosome (lumen) with a certainty of 0.1000. The most likely cleavage site for a NOV8 peptide is between amino acids
 15 15 and 16, at: CTC-SP.

Table 8B. Encoded NOV8 protein sequence (SEQ ID NO:22).

MALLYSSLGVVCTCSPSYFGLTGSEPLTILPLTLEPEAAAQAHYKACDRLKLERKQRRMCRRDPGVAETL
 VEAVSMSALECQFQFRFERWNCTLEGRYRASLLKRGFKETAFLYAISSAGLTHALAKACSAGRMRCTCDE
 APDLENREGWKWGGCSEDIFFGGMVSREFADARENRPDARSAMNRHNEAGRQVIKAGVETTCCKHGVSGS
 CTVRTCWRQLAPFHEVGKHLKHYESALKVGSTTNEAAGEAGAI SPPRGRASGAGSDPLPRTPELVHLD
 SPSFCLAGRFSPGTAGRRCHREKNCESICCGRHNTQSRVTRPCQCQVRWCCYVECRQCTQREEVYTCKG

The disclosed NOV8 amino acid sequence has 270 of 354 amino acid residues (76%) identical to, and 310 of 354 amino acid residues (87%) similar to, the 354 amino acid residue
 20 ptnr:SWISSPROT-ACC:O42280 protein from *Gallus gallus* (Chicken) (WNT-14 Protein Precursor (1.2e⁻¹⁵¹).

NOV8 is expressed in at least brain. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, the sequence is predicted to be expressed in brain because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AF031168|acc:AF031168.1) a closely related [*Gallus gallus* Wnt-14 protein (Wnt-14) mRNA, complete cds].

NOV8 also has homology to the amino acid sequence shown in the BLASTP data listed in Table 8C.

Table 8C. BLAST results for NOV8					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 15082261 ref NP_03386.1 (NM_003395)	wingless-type MMTV integration site family, member 14 [Homo sapiens]	365	306/340 (90%)	321/340 (94%)	e-167
gi 3915306 sp O42280 WN14 CHICK	WNT-14 PROTEIN PRECURSOR	354	270/357 (75%)	310/357 (86%)	e-142
gi 16303264 dbj BAB70499.1 (AB063483)	WNT14B [Homo sapiens]	357	193/339 (56%)	244/339 (71%)	e-100
gi 7106447 ref NP_033548.1 (NM_009522)	wingless-related MMTV integration site 3A [Mus musculus]	352	141/311 (45%)	179/311 (57%)	2e-62
gi 5821261 dbj BAA83743.1 (AB024080)	Wnt-3a [Gallus gallus]	376	139/311 (44%)	179/311 (56%)	3e-62

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 8D.

Table 8D. Information for the ClustalW proteins

- 1) NOV8 (SEQ ID NO:22)
- 2) gi|15082261|ref|NP_003386.1| (NM_003395) wingless-type MMTV integration site family, member 14 [Homo sapiens] (SEQ ID NO:64)
- 3) gi|3915306|sp|O42280|WN14 CHICK WNT-14 PROTEIN PRECURSOR (SEQ ID NO:63)
- 4) gi|16303264|dbj|BAB70499.1| (AB063483) WNT14B [Homo sapiens] (SEQ ID NO:62)
- 5) gi|7106447|ref|NP_033548.1| (NM_009522) wingless-related MMTV integration site 3A [Mus musculus] (SEQ ID NO:67)
- 6) gi|5821261|dbj|BAA83743.1| (AB024080) Wnt-3a [Gallus gallus] (SEQ ID NO:68)

		10	20	30	40	50
25	NOV8	MAILYSSGCVVCT	CSPSYFGLTGSEPLTILE		
	gi 15082261	MLDGSPLARWLAAAFGLTLLAALR	PSAAYFGLTGSEPLTILE			
	gi 3915306	MAILRALGCLLACTP	RPSAAYFGLTGSEPLTILE			
	gi 16303264	MRPPPALALAGLCCLLALP	AAASVFGLTGRFVLTPEFG			
	gi 7106447	MAPLGYLVLCS	LKQA	LG-SYPTWWS		
30	gi 5821261	MKSFCSEVVAKSR	GLKQWGWCGWTPMGS	SAWKWIS	SEQRSS	DELW
		60	70	80	90	100
35	NOV8	LTLEP-EAAACAHYKACDRLK	LEKKORMCRRDPGVAETLVEAVS		
	gi 15082261	LTLEP-EAAACAHYKACDRLK	LEKKORMCRRDPGVAETLVEAVS			
	gi 3915306	LTSEMEDAAVKAHYKACDRLK	LEKKORMCRRDPGCAETLMEATIS			
	gi 16303264	LGTAAPAPQGGAHLCDCDLK	LSRRKQLCRRFPGLAETLRDAAH			
	gi 7106447	LAVGPOVSSLSLTPIL	CASTPGIVPQGLF	PCPNYVEIMPSVAFGVK		

gi 5821261		DVGS	LAIGHQYSSLGTPILGSGTIPGIVPKQLRFRNYVEIMPVAGGVK			
		110	120	130	140	150
5	NOV8	MSALECCQ	QFRERWNC	TLEGR---	YFASLLKRGFKETAFLYA	ISSAGLT
	gi 15082261	MSALECCQ	QFRERWNC	TLEGR---	YFASLLKRGFKETAFLYA	ISSAGLT
	gi 3915306	MSALECCQ	QFRERWNC	TLEGR---	YFASLLKRGFKETAFLYA	ISSAGLT
	gi 16303264	LGLLECCQ	QFRERWNC	SLLEG---	FTGLLKRGFKETAFLYA	VSSAALT
10	gi 7106447	AGIQECQ	QFRGRWNC	TVSNSLAIFGPVLDKATRESAFVHATA	SAGVA	
	gi 5821261	LGIQECQ	QFRGRWNC	TVNDSLAIIFGPVLDKATRESAFVHATA	SAGVA	
		160	170	180	190	200
15	NOV8	HALAKACS	AGRMERCTCDEAPDLENRE	GWGGGSE	EDIEFGGMVSR	REHAD
	gi 15082261	HALAKACS	AGRMERCTCDEAPDLENRE	AWGGGCGDNLKYSSKFVREFLG		
	gi 3915306	HAMAKACS	AGRMERCTCDEAPDLENRE	AWGGGCGDNLKYSSKFVREFLG		
	gi 16303264	HTLAKACS	AGRMERCTCDDSPGLE	SRQAHGVC	GDNLKYSTKFLSNELG	
	gi 7106447	FAVTRSCA	EGSAAICGCSRLQ	GSPEGGWGGGSE	EDIEFGGMVSR	REHAD
20	gi 5821261	FAVTRSCA	EGSATICGCDTRHKG	SPGEGGWGGGSE	EDIEFGGMVSR	REHAD
		210	220	230	240	250
25	NOV8	ARENRPD	ASAMNRHNNEAGRQVIRAGV	ETTCCKCHGVSGSCTV	RTCWRL	
	gi 15082261	RPSRK-	DLRARVDFHNNLVGVKVIKAGV	ETTCCKCHGVSGSCTV	RTCWRL	
	gi 3915306	RKPNK-	DLRARVDFHNNLVGMKVIKAGV	ETTCCKCHGVSGSCTV	RTCWRL	
	gi 16303264	SKRGNKDL	RRADAHNTHVGIKAVESCLR	TTCKCHGVSGSCAV	RTCWRL	
	gi 7106447	ARENRPD	ASAMNRHNNEAGRQATASHMHLKCKCHG	LSGSCEVKT	CWWSQ	
	gi 5821261	ARENRPD	ASAMNRHNNEAGRTSIIELMHLKCKCHG	LSGSCEVKT	CWWSQ	
		260	270	280	290	300
30	NOV8	APFHEVG	KHLRKYETSLKVGSTTNEAAGEAGAI	SPPRGRASGAGGSDP		
	gi 15082261	APFHEVG	KHLRKYETSLKVGSTTNEAAGEAGAI	SPPRGRASGAGGSDP		
	gi 3915306	SPFHEIG	KOLNCKYETSLKVGSTTNEATGE	-GDISPEK--	KSIPGHSDQI	
35	gi 16303264	SPFRETS	QVLLRLYDSAVKVSATNEALRLLELWAPAF	--Q-GSLTKGLA		
	gi 7106447	PDERTIG	DFLELDYDSASEMVVEKHRESRGWVETLRPP	--Y----	TYFKY	
	gi 5821261	PDERVTL	DOYLEDYDSASEMVVEKHRESRGWVETLRPK	--Y----	NFFKA	
		310	320	330	340	350
40	NOV8	PPTPELV	HLDDSPSFCLAG--	RFSPGTAGRRCHREKN--	CESTICCGRG	
	gi 15082261	PPTPELV	HLDDSPSFCLAG--	RFSPGTAGRRCHREKN--	CESTICCGRG	
	gi 3915306	PRTDLV	YIDDSPSFCLMS--	RYSPGTSGRKCYKDKN--	CDSTICCGRG	
	gi 16303264	PRSGDLV	YMEDSPSFCLPS--	KYSPGTAGRVCSREAS--	CSSLCCGRG	
45	gi 7106447	PTERDLV	YVYASPNFCEPNPETGSEF	GTRDRICNVSSHGIDG	CDLLCCGRG	
	gi 5821261	PTERDLV	YVYASPNFCEPNPETGSEF	GTRDRICNVSSHGIDG	CDLLCCGRG	
		360	370	380		
50	NOV8	HNTQSRV	VTRPCCQVRWCCYVECRQCTCR	EEVYTCRG		
	gi 15082261	HNTQSPV	VTRPCCQVRWCCYVECRQCTCR	EEVYTCRG		
	gi 3915306	HNTQSRV	VTRPCCQVRWCCYVECRQCTCR	EEVYTCRD		
	gi 16303264	YDTQSEL	VAFSCHCQVQWCCYVECCQCV	CELLVYTCRH		
	gi 7106447	HNARTERR	ERKCHVFWCCYVSCQECTRV	YDVHTCH		
55	gi 5821261	HNTETEK	RKERKCHLFHWCCYVRCQECTRV	YDVHTCH		

Tables 8E and 8F list the domain descriptions from DOMAIN analysis results against NOV8. This indicates that the NOV8 sequence has properties similar to those of other proteins known to contain this domain.

Table 8E. Domain Analysis of NOV8

gnl|Pfam|pfam00110, wnt, wnt family. (SEQ ID NO:104)
 CD-Length = 313 residues, 99.7% aligned
 Score = 313 bits (801), Expect = 1e-86

20

gnl|Smart|smart00097, WNT1, found in Wnt-1 (SEQ ID NO:105)
CD-Length = 304 residues, 98.7% aligned
Score = 292 bits (748), Expect = 2e-80

25

50

and null mutant mice. Wnts are involved in processes involved in mammary gland development and cancer. Recent studies have demonstrated that these molecules are critical to organogenesis of several systems, such as the kidney and brain. Wnts regulate the early development, i.e. neural induction, and their role persists in later stages of development as well as in the mature organ. An example of this is seen in the brain, where the loss of certain Wnts leads to the absence of critical regions of the brain, e.g. the hippocampus, involved in learning and memory, or the cerebellum, involved in motor function. Wnts have also been implicated in the genesis of degenerative diseases such as Alzheimer's disease. The protein encoded by the novel gene described herein may therefore play a role in cellular proliferation, differentiation, dysregulation, organogenesis and disease processes such as cancer, developmental defects etc.

Alzheimer's disease (AD) is a neurodegenerative disease with progressive dementia accompanied by three main structural changes in the brain: diffuse loss of neurons; intracellular protein deposits termed neurofibrillary tangles (NFT) and extracellular protein deposits termed amyloid or senile plaques, surrounded by dystrophic neurites. Two major hypotheses have been proposed in order to explain the molecular hallmarks of the disease: The 'amyloid cascade' hypothesis and the 'neuronal cytoskeletal degeneration' hypothesis. While the former is supported by genetic studies of the early-onset familial forms of AD (FAD), the latter revolves around the observation in vivo that cytoskeletal changes - including the abnormal phosphorylation state of the microtubule associated protein tau - may precede the deposition of senile plaques. Recent studies have suggested that the trafficking process of membrane associated proteins is modulated by the FAD-linked presenilin (PS) proteins, and that amyloid beta-peptide deposition may be initiated intracellularly, through the secretory pathway. Current hypotheses concerning presenilin function are based upon its cellular localization and its putative interaction as macromolecular complexes with the cell-adhesion/signaling beta-catenin molecule and the glycogen synthase kinase 3beta (GSK-3beta) enzyme. Developmental studies have shown that PS proteins function as components in the Notch signal transduction cascade and that beta-catenin and GSK-3beta are transducers of the Wnt signaling pathway. Both pathways are thought to have an important role in brain development, and they have been connected through Dishevelled (Dvl) protein, a known transducer of the Wnt pathway.

Members of the vertebrate Wnt family have been subdivided into two functional classes according to their biological activities. Some Wnts signal through the canonical Wnt-1/wingless pathway by stabilizing cytoplasmic beta-catenin. By contrast other Wnts stimulate

intracellular Ca²⁺ release and activate two kinases, CamKII and PKC, in a G-protein-dependent manner. Moreover, putative Wnt receptors belonging to the Frizzled gene family have been identified that preferentially couple to the two prospective pathways in the absence of ectopic Wnt ligand and that might account for the signaling specificity of the Wnt pathways. As Ca²⁺ release was the first described feature of the noncanonical pathway, and as Ca²⁺ probably plays a key role in the activation of CamKII and PKC, Kuhl M, et al., (*Trends Genet* 2000 Jul;16(7):279-83) have named this Wnt pathway the Wnt/Ca²⁺ pathway.

Many constituents of Wnt signaling pathways are expressed in the developing and mature nervous systems. Recent work has shown that Wnt signaling controls initial formation of the neural plate and many subsequent patterning decisions in the embryonic nervous system, including formation of the neural crest. Wnt signaling continues to be important at later stages of development. Wnts have been shown to regulate the anatomy of the neuronal cytoskeleton and the differentiation of synapses in the cerebellum. Wnt signaling has been demonstrated to regulate apoptosis and may participate in degenerative processes leading to cell death in the aging brain.

Recent genetic studies have shown that the signalling factor Wnt3a is required for formation of the hippocampus; the developmental consequences of Wnt signalling in the hippocampus are mediated by multiple HMG-box transcription factors, with LEF-1 being required just for formation of the dentate gyrus.

Wnt-1 was first identified as a protooncogene activated by viral insertion in mouse mammary tumors. Transgenic expression of this gene using a mouse mammary tumor virus LTR enhancer causes extensive ductal hyperplasia early in life and mammary adenocarcinomas in approximately 50% of the female transgenic (TG) mice by 6 months of age. Metastasis to the lung and proximal lymph nodes is rare at the time tumors are detected but frequent after the removal of the primary neoplasm. The potent mitogenic effect mediated by Wnt-1 expression does not require estrogen stimulation; tumors form after an increased latency in estrogen receptor alpha-null mice. Several genetic lesions, including inactivation of p53 and over-expression of Fgf-3, collaborate with Wnt-1 in leading to mammary tumors, but loss of Sky and inactivation of one allele of Rb do not affect the rate of tumor formation in Wnt-1 TG mice.

Communication between cells is often mediated by secreted signaling molecules that bind cell surface receptors and modulate the activity of specific intracellular effectors. The Wnt family of secreted glycoproteins is one group of signaling molecules that has been shown to control a variety of developmental processes including cell fate specification, cell

proliferation, cell polarity and cell migration. In addition, mis-regulation of Wnt signaling can cause developmental defects and is implicated in the genesis of several human cancers. The importance of Wnt signaling in development and in clinical pathologies is underscored by the large number of primary research papers examining various aspects of Wnt signaling that have been published in the past several years.

Reproductive tract development and function is regulated by circulating steroid hormones. In the mammalian female reproductive tract, estrogenic compounds direct many aspects of cytodifferentiation including uterine gland formation, smooth muscle morphology, and epithelial differentiation. While it is clear that these hormones act through their cognate nuclear receptors, it is less clear what signaling events follow hormonal stimulation that govern cytodifferentiation. Recent advances in molecular embryology and cancer cell biology have identified the Wnt family of secreted signaling molecules. Discussed here are recent advances that point to a definitive role during uterine development and adult function for one member of the Wnt gene family, Wnt-7a. In addition, recent data is reviewed that implicates Wnt-7a deregulation in response to pre-natal exposure to the synthetic estrogenic compound, DES. These advances point to an important role for the Wnt gene family in various reproductive tract pathologies including cancer.

Holoprosencephaly (HPE) is the most common developmental defect of the forebrain in humans. Several distinct human genes for holoprosencephaly have now been identified. They include Sonic hedgehog (SHH), ZIC2, and SIX3. Many additional genes involved in forebrain development are rapidly being cloned and characterized in model vertebrate organisms. These include Patched (Ptc), Smoothened (Smo), cubitus interruptus (ci)/Gli, wingless (wg/Wnt, decapentaplegic (dpp)/BMP, Hedgehog interacting protein (Hip), nodal, Smads, One-eyed pinhead (Oep), and TG-Interacting Factor (TGIF). However, further analysis is needed before their roles in HPE can be established.

Female reproductive hormones control mammary gland morphogenesis. In the absence of the progesterone receptor (PR) from the mammary epithelium, ductal side-branching fails to occur. Briskin C, et al. (*Genes Dev* 2000 Mar 15;14(6):650-4) overcame this defect by ectopic expression of the protooncogene Wnt-1. Transplantation of mammary epithelia from Wnt-4(-)/(-) mice shows that Wnt-4 has an essential role in side-branching early in pregnancy. PR and Wnt-4 mRNAs colocalize to the luminal compartment of the ductal epithelium. Progesterone induces Wnt-4 in mammary epithelial cells and is required for increased Wnt-4 expression during pregnancy. Thus, Wnt signaling is essential in mediating progesterone function during mammary gland morphogenesis.

Synapse formation requires changes in cell morphology and the upregulation and localization of synaptic proteins. In the cerebellum, mossy fibers undergo extensive remodeling as they contact several granule cells and form complex, multisynaptic glomerular rosettes. Hall AC, et al., (*Cell* 2000 Mar 3;100(5):525-35) showed that granule cells secrete factors that induce axon and growth cone remodeling in mossy fibers. This effect is blocked by the WNT antagonist, sFRP-1, and mimicked by WNT-7a, which is expressed by granule cells. WNT-7a also induces synapsin I clustering at remodeled areas of mossy fibers, a preliminary step in synaptogenesis. Wnt-7a mutant mice show a delay in the morphological maturation of glomerular rosettes and in the accumulation of synapsin I. We propose that WNT-7a can function as a synaptogenic factor.

Estrogens have important functions in mammary gland development and carcinogenesis. To better define these roles, Bocchinfuso WP, et al., (*Cancer Res* 1999 Apr 15;59(8):1869-76) have used two previously characterized lines of genetically altered mice: estrogen receptor-alpha (ER alpha) knockout (ERKO) mice, which lack the gene encoding ER alpha, and mouse mammary virus tumor (MMTV)-Wnt-1 transgenic mice (Wnt-1 TG), which develop mammary hyperplasia and neoplasia due to ectopic production of the Wnt-1 secretory glycoprotein. Bocchinfuso WP, et al. have crossed these lines to ascertain the effects of ER alpha deficiency on mammary gland development and carcinogenesis in mice expressing the Wnt-1 transgene. Introduction of the Wnt-1 transgene into the ERKO background stimulates proliferation of alveolar-like epithelium, indicating that Wnt-1 protein can promote mitogenesis in the absence of an ER alpha-mediated response. The hyperplastic glandular tissue remains confined to the nipple region, implying that the requirement for ER alpha in ductal expansion is not overcome by ectopic Wnt-1. Tumors were detected in virgin ERKO females expressing the Wnt-1 transgene at an average age (48 weeks) that is twice that seen in virgin Wnt-1 TG mice (24 weeks) competent to produce ER alpha. Prepubertal ovariectomy of Wnt-1 TG mice also extended tumor latency to 42 weeks. However, pregnancy did not appear to accelerate the appearance of tumors in Wnt-1 TG mice, and tumor growth rates were not measurably affected by late ovariectomy. Small hyperplastic mammary glands were observed in Wnt-1 TG males, regardless of ER alpha gene status; the glands were similar in appearance to those found in ERKO/Wnt-1 TG females. Mammary tumors also occurred in Wnt-1 TG males; latency tended to be longer in the heterozygous ER alpha and ERKO males (86 to 100 weeks) than in wild-type ER alpha mice (ca. 75 weeks). Bocchinfuso WP, et al. concluded that ectopic expression of the Wnt-1 proto-oncogene can induce mammary hyperplasia and tumorigenesis in the absence of ER alpha in female and male mice. The delayed time of tumor

appearance may depend on the number of cells at risk of secondary events in the hyperplastic glands, on the carcinogenesis-promoting effects of ER alpha signaling, or on both.

Wnt-1 and Wnt-3a proto-oncogenes have been implicated in the development of midbrain and hindbrain structures. Evidence for such a role has been derived from in situ hybridization studies showing Wnt-1 and -3a expression in developing cranial and spinal cord regions and from studies of mutant mice whose Wnt-1 genes have undergone targeted disruption by homologous recombination. Wnt-1 null mutants exhibit cranial defects but no spinal cord abnormalities, despite expression of the gene in these regions. The absence of spinal cord abnormalities is thought to be due to a functional compensation of the Wnt-1 deficiency by related genes, a problem that has complicated the analysis of null mutants of other developmental genes as well. Augustine K, et al., (*Dev Genet* 1993;14(6):500-20) describe the attenuation of Wnt-1 expression using antisense oligonucleotide inhibition in mouse embryos grown in culture. Augustine K, et al. induced similar mid- and hindbrain abnormalities as those seen in the Wnt-1 null mutant mice. Attenuation of Wnt-1 expression was also associated with cardiomegaly resulting in hemostasis. These findings are consistent with the possibility that a subset of Wnt-1 expressing cells include neural crest cells known to contribute to septation of the truncus arteriosus and to formation of the visceral arches. Antisense knockout of Wnt-3a, a gene structurally related to Wnt-1, targeted the forebrain and midbrain region, which were hypoplastic and failed to expand, and the spinal cord, which exhibited lateral outpocketings at the level of the forelimb buds. Dual antisense knockouts of Wnt-1 and Wnt-3a targeted all brain regions leading to incomplete closure of the cranial neural folds, and an increase in the number and severity of outpocketings along the spinal cord, suggesting that these genes complement one another to produce normal patterning of the spinal cord. The short time required to assess the mutant phenotype (2 days) and the need for limited sequence information of the target gene (20-25 nucleotides) make this antisense oligonucleotide/whole embryo culture system ideal for testing the importance of specific genes and their interactions in murine embryonic development.

Wnt-1 (previously known as int-1) is a proto-oncogene induced by the integration of the mouse mammary tumor virus. It is thought to play a role in intercellular communication and seems to be a signalling molecule important in the development of the central nervous system (CNS). The sequence of wnt-1 is highly conserved in mammals, fish, and amphibians. Wnt-1 is a member of a large family of related proteins that are all thought to be developmental regulators. These proteins are known as wnt-2 (also known as irp), wnt-3 up to wnt-15. At least four members of this family are present in *Drosophila*. One of them, wingless

(wg), is implicated in segmentation polarity. All these proteins share the following features characteristics of secretory proteins, a signal peptide, several potential N-glycosylation sites and 22 conserved cysteines that are probably involved in disulfide bonds. The Wnt proteins seem to adhere to the plasma membrane of the secreting cells and are therefore likely to signal over only few cell diameters.

The disclosed NOV8 nucleic acid of the invention encoding a Wnt-14-like protein includes the nucleic acid whose sequence is provided in Table 8A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 8A while still encoding a protein that maintains its Wnt-14-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 23 percent of the bases may be so changed.

The disclosed NOV8 protein of the invention includes the Wnt-14-like protein whose sequence is provided in Table 8B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 8B while still encoding a protein that maintains its Wnt-14-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 24 percent of the residues may be so changed.

The protein similarity information, expression pattern, and map location for the Wnt-14-like protein and nucleic acid (NOV8) disclosed herein suggest that NOV8 may have important structural and/or physiological functions characteristic of the Wnt-14-like family. Therefore, the NOV8 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic

applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo.

5 The NOV8 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, 10 Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, cancer, developmental defects, and/or other pathologies/disorders. The NOV8 nucleic acid, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

15 NOV8 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV8 protein have multiple hydrophilic regions, 20 each of which can be used as an immunogen. In one embodiment, contemplated NOV8 epitope is from about amino acids 40 to 70. In another embodiment, the contemplated NOV8 epitope is from about amino acids 80 to 110. In further embodiments, the contemplated NOV8 epitope is from about amino acids 120 to 200, from about amino acids 220 to 245, from about amino acids 250 to 280, or from about amino acids 290 to 340. This novel protein also has 25 value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV9

30 A disclosed NOV9 nucleic acid of 2037 nucleotides (also referred to as AC069250_28_da1) encoding a beta-adrenergic receptor kinase-like protein is shown in Table 9A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 16-18 and ending with a TGA codon at nucleotides 2020-2022. A putative untranslated region upstream from the initiation codon and downstream from the termination

codon is underlined in Table 9A. The start and stop codons are in bold letters. Single nucleotide polymorphism data for NOV9 is discussed in further detail in Example 3.

Table 9A. NOV9 nucleotide sequence (SEQ ID NO:23).

GCCGCCGCCGCCAAGATGGCGGACCTGGAGGCGGTGCTGGCCGACGTGAGCTACCTGATGGCCATGGAGAAG
AGCAAGGCCACGCCGCCGCCGCCAGCAAGAAGATACTGCTGCCCGAGCCAGCATCCGCGAGTGTCTATG
CAGAAGTACCTGGAGGACCGGGGCGAGGTGACCTTTGAGAAGATCTTTCCAGAAAGCTGGGGTACCTGCTC
TTCCGAGACTTCTGCTGAACACCTGGAGGAGGCCAGGCCCTTGGTGAATTCTATGAGGAGATCAAGAAG
TACGAGAAGCTGGAGACGGAGGAGGAGCGTGTGGCCCGCAGCCGGGAGATCTCGACTCATACATCATGAAG
GAGCTGCTGGCCTGCTCGCATCCCTTCTCGAAGAGTGCCACTGAGCATGTCCAAGGCCACCTGGGGAAGAAG
CAGGTGCCCTCCGGATCTCTTCCAGCCATACATCGAAGAGATTGTCAAAACCTCCGAGGGGACGTGTTCCAG
AAATTCATTGAGAGCGATAAGTTCACACGGTTTTGCCAGTGGAGAATGTGGAGCTCAACATCCACCTGACC
ATGAATGACTTCAGCGTGCATCGCATCATTGGGCGCGGGGGCTTTGGCGAGGTCTATGGGTGCCGGAAGGCT
GACACAGGCAAGATGTACGCCATGAAGTGCCTGGACAAAAGCGCATCAAGATGAAGCAGGGGAGACCCTG
GCCCTGAACGAGCGCATCATGCTCTCGCTCGTCAGCACTGGGGAATGCCCATTCATTGTCTGCATGTCTATC
GCGTTCACACGCCAGACAAGCTCAGCTTCATCCTGGACCTCATGAACGGTGGGGACCTGCACTACCACTC
TCCAGCACGGGGTCTTCTCAGAGGCTGACATGCGCTTCTATGCGGCGAGATCATCCTGGGCTCGAGCAC
ATGCACAACCGCTTCGTGGTCTACCGGGACCTGAAGCCAGCCAACATCCTTCTGGACGAGCATGGCCACGTG
CGGATCTCGGACCTGGGCTGGCCTGTGACTTCTCCAAGAAGAAGCCCATGCCAGCGTGGGCACCCACGGG
TACATGGCTCCGGAGGTCTGCAGAAGGGCGTGGCCTACGACAGCAGTGGCGACTGGTTCTCTCTGGGGTGC
ATGCTCTTCAAGTTCGTGCGGGGCGACAGCCCTTCCGGCAGCACAAAGACCAAGACAAGCATGAGATCGAC
CGCATGACGCTGACGATGGCCGTGGAGCTGCCGACTCCTTCTCCCTGAACCTACGCTCCCTGCTGGAGGGG
TTGCTGCAGAGGGATGTCAACCGAGATTGGGCTGCCTGGGCGAGGGGCTCAGGAGGTGAAGAGAGCCCC
TTTTTCCGCTCCCTGGACTGGCAGATGGTCTTCTTGCAGAAGTACCTCCCCGCTGATCCCCCACGAGGG
GAGGTGAACGCGGCGACGCTTCGACATTGGCTCCTTCGATGAGGAGGACAAAAGGAATCAAGCAGGAG
GTGGCAGAGACTGTCTTCGACACCATCAACGCTGAGACAGACCGGCTGGAGGCTCGCAAGAAAGCCAGAAG
AAGCAGCTGGGCCATGAGGAAGACTACGCCCTGGGCAAGGACTGCATCATGCATGGCTACATGTCCAAGATG
GGCAACCCCTTCTTGACCCAGTGGCAGCGCGGTACTTCTACCTGTTCCCAACCGCTCGAGTGGCGGGG
GAGGGCGAGGCCCCGAGAGCTGCTGACCATGGAGGAGATCCAGTGGTGGAGGAGACGCAGATCAAGGAG
CGCAAGTGCCTGCTCTCAAGATCCGCGGTGGGAACAGTTCATTTGCAAGTGCATAGCGACCTGAGCTG
GTGCAAGTGAAGAAGGAGCTGCGCGACGCTACCGCGAGGCCAGCAGCTGGTGCAGCGGGTGGCCAAAGATG
AAGAACAAGCCGCGCTCGCCGCTGGTGGAGCTGAGCAAGTGGCGCTGGTCCAGCGCGCAGTGCCCAACGGC
CTCTGACCCGCCCCACCCGCT

5 In a search of public sequence databases, the NOV9 nucleic acid sequence, located on chromosome 11 has 1546 of 1574 bases (98%) identical to a beta-adrenergic receptor kinase 1 mRNA from *Homo sapiens*, (GENBANK-ID: HUMBARK1A) (E = 0.0). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

10 The disclosed NOV9 polypeptide (SEQ ID NO:24) encoded by SEQ ID NO:23 has 668 amino acid residues and is presented in Table 9B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV9 has no signal peptide and is likely to be localized in the nucleus with a certainty of 0.3000. In other embodiments, NOV9 may also be localized to the microbody (peroxisome) with a certainty of 0.1478, the mitochondrial matrix (lumen) with a certainty of 0.1000 or in the lysosome (lumen) with a
15 certainty of 0.1000.

Table 9B. Encoded NOV9 protein sequence (SEQ ID NO:24).

MADLEAVLADVSYLMAMEKSKATPAARASKKILLPEPSIRSVMQYLEDRGEVTFEKIFSQKLGYYLLFRDFC
LNHLEEARPLVEFYEEIKKYEKLETEEERVAREREIFDSYIMKELLACSHPFSSKATEHVQGHGKQVPPD

```

LFQPYIEEICQNLRGDVFOKFIESDKFTRFCQWKNVELNIHLTMNDFS VHRITIGRGGFGEVYGCRAKADTKGM
YAMKCLDKKRIKMKQGETLALNERIMLSLVSTGDCPFIVCMSYAFHTPDKLSFILDLMNGGDLHYHLSQHG
V FSEADMRFYAAEII LGLEHMHNRFVYRDLK PANILLDEHGHVRISDLGLACDFS KKKPHASVGT HGYMAPE
VLQKGVAYDSSADWFS LGCMLEKLLRGHS PFRQHKTKDKHEIDRMTLTMAVELPDSFSPELRSLLLEGLLQ
RD VNRRLGCLGRGAQEVKES PFFRS LDWQMVFLQKYPPLIPRGEVNAAADAFDIGSFDEEDTKGKIQEVAETV
FDTINAETDRLEARKKAKNKQLGHEEDYALGKDCIMHGYMSKMGNPFLTQWQRRYFYLFNRLLEWRGEGEAP
QSLTMEELIQSVEETQIKERKCLLLKIRGGKQFILQCDSDPELVQWKELRDAYREAQQLVQRPVKMKNKPR
SPVVELSKVPLVQSGSANG

```

A search of sequence databases reveals that the NOV9 amino acid sequence has 495 of 497 amino acid residues (99%) identical to, and 495 of 497 amino acid residues (99%) similar to, the 689 amino acid residue beta-adrenergic receptor kinase from *Homo sapiens* (A53791) (E = 0.0). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV9 is expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain - whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, the sequence is predicted to be expressed in blood leukocytes because of the expression pattern of (GENBANK-ID:gb:GENBANK-ID:HUMBARK1A|acc:M80776.1) a closely related Human beta-adrenergic receptor kinase 1 mRNA, complete cds homolog in species *Homo sapiens*.

The disclosed NOV9 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 9C.

Table 9C. BLAST results for NOV9					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
ptnr:pir-id:A53791	beta-adrenergic- receptor kinase (EC 2.7.1.126) 1 - human	689	495/497 (99%)	495/497 (99%)	0.0
ptnr:SWISSPROT- ACC:P25098	Beta-adrenergic receptor kinase 1 (EC 2.7.1.126)	689	494/497 (99%)	495/497 (99%)	0.0
ptnr:SPTREMBL- ACC:Q99LL8	SIMILAR TO ADRENERGIC, BETA, RECEPTOR KINASE 1 - <i>Mus musculus</i>	687	490/495 (98%),	493/495 (99%)	0.0

ptnr:SWISSPROT- ACC:P26817	Beta-adrenergic receptor kinase 1	689	489/497 (98%)	493/497 (99%)	0.0
ptnr:SPTREMBL- ACC:Q99MK8	G PROTEIN RECEPTOR KINASE 2	689	490/497 (98%)	494/497 (99%)	0.0

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 9D. In the ClustalW alignment of the NOV9 proteins, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 9D. ClustalW Analysis of NOV9

- 1) NOV9 (SEQ ID NO:24)
- 2) ptnr:pir-id:A53791 beta-adrenergic-receptor kinase (EC 2.7.1.126) 1 (SEQ ID NO:69)
- 3) ptnr:SWISSPROT-ACC:P25098 Beta-adrenergic receptor kinase 1 (EC 2.7.1.126) (SEQ ID NO:70)
- 4) ptnr:SPTREMBL-ACC:Q99LL8 SIMILAR TO ADRENERGIC, BETA, RECEPTOR KINASE 1 - Mus musculus (Mouse) (SEQ ID NO:71)
- 5) 6) ptnr:SWISSPROT-ACC:P26817 Beta-adrenergic receptor kinase 1 (EC 2.7.1.126) (Beta-ARK-1) (SEQ ID NO:72)

NOV9	MADLEAVLADVSYLMAMEKSKATPAARASKKILLPEPSIRSVMQKYLEDRGEVTFEKIFS	60
A53791	MADLEAVLADVSYLMAMEKSKATPAARASKKILLPEPSIRSVMQKYLEDRGEVTFEKIFS	60
P25098	MADLEAVLADVSYLMAMEKSKATPAARASKKILLPEPSIRSVMQKYLEDRGEVTFEKIFS	60
Q99LL8	--DLEAVLADVSYLMAMEKSKATPAARASKKILLPEPSIRSVMQKYLEDRGEVTFEKIFS	58
P26817	MADLEAVLADVSYLMAMEKSKATPAARASKKILLPEPSIRSVMQKYLEDRGEVTFEKIFS	60
NOV9	QKLGYYLFRDFCLNHLEEARPLVEFYEEIKKYEKLETEEERVARSRIFDSYIMKELLAC	120
A53791	QKLGYYLFRDFCLNHLEEARPLVEFYEEIKKYEKLETEEERVARSRIFDSYIMKELLAC	120
P25098	QKLGYYLFRDFCLNHLEEARPLVEFYEEIKKYEKLETEEERVARSRIFDSYIMKELLAC	120
Q99LL8	QKLGYYLFRDFCLNHLEEARPLVEFYEEIKKYEKLETEEERVARSRIFDSYIMKELLAC	118
P26817	QKLGYYLFRDFCLNHLEEARPLVEFYEEIKKYEKLETEEERVARSRIFDSYIMKELLAC	120
NOV9	SHPFKSATEHVQGHGKQVPPDLFQPYIEEIQNLRGDVFQKFIESDKFTRFCQWKNV	180
A53791	SHPFKSATEHVQGHGKQVPPDLFQPYIEEIQNLRGDVFQKFIESDKFTRFCQWKNV	180
P25098	SHPFKSATEHVQGHGKQVPPDLFQPYIEEIQNLRGDVFQKFIESDKFTRFCQWKNV	180
Q99LL8	SHPFKSATEHVQGHGKQVPPDLFQPYIEEIQNLRGDVFQKFIESDKFTRFCQWKNV	178
P26817	SHPFKSATEHVQGHGKQVPPDLFQPYIEEIQNLRGDVFQKFIESDKFTRFCQWKNV	180
NOV9	ELNIHLTMNDFSVHRIIGRGGFGEVYGCRRKADTGKMYAMKCLDKKRIKMKQGETLALNER	240
A53791	ELNIHLTMNDFSVHRIIGRGGFGEVYGCRRKADTGKMYAMKCLDKKRIKMKQGETLALNER	240
P25098	ELNIHLTMNDFSVHRIIGRGGFGEVYGCRRKADTGKMYAMKCLDKKRIKMKQGETLALNER	240
Q99LL8	ELNIHLTMNDFSVHRIIGRGGFGEVYGCRRKADTGKMYAMKCLDKKRIKMKQGETLALNER	238
P26817	ELNIHLTMNDFSVHRIIGRGGFGEVYGCRRKADTGKMYAMKCLDKKRIKMKQGETLALNER	240
NOV9	IMLSLVSTGDCPFIVCMASYAFHTPKLSFILDLMNGGDLHYHLSQHGVFSEADMRFYAAE	300
A53791	IMLSLVSTGDCPFIVCMASYAFHTPKLSFILDLMNGGDLHYHLSQHGVFSEADMRFYAAE	300
P25098	IMLSLVSTGDCPFIVCMASYAFHTPKLSFILDLMNGGDLHYHLSQHGVFSEADMRFYAAE	300
Q99LL8	IMLSLVSTGDCPFIVCMASYAFHTPKLSFILDLMNGGDLHYHLSQHGVFSEADMRFYAAE	298
P26817	IMLSLVSTGDCPFIVCMASYAFHTPKLSFILDLMNGGDLHYHLSQHGVFSEADMRFYAAE	300
NOV9	IILGLEHMHNRFFVYRDLKPANILLDEHGHVRISDLGLACDFSKKKPHASVGTGHYMAPE	360
A53791	IILGLEHMHNRFFVYRDLKPANILLDEHGHVRISDLGLACDFSKKKPHASVGTGHYMAPE	360

	P25098	IILGLEHMHNR	RVVYRDLKPAN	IILLDEHGHVRISDLGLACDFS	SKKPHASVGT	HGYMAPE	360																																										
	Q99LL8	IILGLEHMHNR	RVVYRDLKPAN	IILLDEHGHVRISDLGLACDFS	SKKPHASVGT	HGYMAPE	358																																										
	P26817	IILGLEHMHNR	RVVYRDLKPAN	IILLDEHGHVRISDLGLACDFS	SKKPHASVGT	HGYMAPE	360																																										
5	NOV9	VLQKGVA	YDSSADW	FLGCMLE	FKLLRGH	SPFRQHKTKDKHEIDRMTLTMAVELPDSFSPE	420																																										
	A53791	VLQKGVA	YDSSADW	FLGCMLE	FKLLRGH	SPFRQHKTKDKHEIDRMTLTMAVELPDSFSPE	420																																										
	P25098	VLQKGVA	YDSSADW	FLGCMLE	FKLLRGH	SPFRQHKTKDKHEIDRMTLTMAVELPDSFSPE	420																																										
	Q99LL8	VLQKGVA	YDSSADW	FLGCMLE	FKLLRGH	SPFRQHKTKDKHEIDRMTLTMAVELPDSFSPE	418																																										
	P26817	VLQKGVA	YDSSADW	FLGCMLE	FKLLRGH	SPFRQHKTKDKHEIDRMTLTMAVELPDSFSPE	420																																										
10	NOV9	LRSLL	EGLLQ	RDVNR	RRLGCL	GRGAQEVKESPFERSLDWQMVFLQKYPPLIPPRGEVNAA	480																																										
	A53791	LRSLL	EGLLQ	RDVNR	RRLGCL	GRGAQEVKESPFERSLDWQMVFLQKYPPLIPPRGEVNAA	480																																										
	P25098	LRSLL	EGLLQ	RDVNR	RRLGCL	GRGAQEVKESPFERSLDWQMVFLQKYPPLIPPRGEVNAA	480																																										
	Q99LL8	LRSLL	EGLLQ	RDVNR	RRLGCL	GRGAQEVKESPFERSLDWQMVFLQKYPPLIPPRGEVNAA	478																																										
15	P26817	LRSLL	EGLLQ	RDVNR	RRLGCL	GRGAQEVKESPFERSLDWQMVFLQKYPPLIPPRGEVNAA	480																																										
	NOV9	DAFDIGS	FDEEDTKGIK	-----	Q	EVAETV	FD	TINAE	DRLEARK	519																																							
	A53791	DAFDIGS	FDEEDTKGIK	LLDS	QOELYN	FP	LTISER	WQ	Q	EVAETV	FD	TINAE	DRLEARK	540																																			
	P25098	DAFDIGS	FDEEDTKGIK	LLDS	QOELYN	FP	LTISER	WQ	Q	EVAETV	FD	TINAE	DRLEARK	540																																			
20	Q99LL8	DAFDIGS	FDEEDTKGIK	LLDS	QOELYN	FP	LTISER	WQ	Q	EVAETV	FD	TINAE	DRLEARK	538																																			
	P26817	DAFDIGS	FDEEDTKGIK	LLDS	QOELYN	FP	LTISER	WQ	Q	EVAETV	FD	TINAE	DRLEARK	540																																			
	NOV9	KAKNK	Q	LGHEED	YALGK	DCIM	HGYMSKMG	NP	FLTQ	W	Q	RRY	F	Y	L	P	N	R	L	E	W	R	G	E	G	E	A	P	Q	S	L	579																	
	A53791	KAKNK	Q	LGHEED	YALGK	DCIM	HGYMSKMG	NP	FLTQ	W	Q	RRY	F	Y	L	P	N	R	L	E	W	R	G	E	G	E	A	P	Q	S	L	600																	
25	P25098	KAKNK	Q	LGHEED	YALGK	DCIM	HGYMSKMG	NP	FLTQ	W	Q	RRY	F	Y	L	P	N	R	L	E	W	R	G	E	G	E	A	P	Q	S	L	600																	
	Q99LL8	KAKNK	Q	LGHEED	YALGK	DCI	V	HGYMSKMG	NP	FLTQ	W	Q	RRY	F	Y	L	P	N	R	L	E	W	R	G	E	G	E	A	P	Q	S	L	598																
	P26817	KAKNK	Q	LGHEED	YALGK	DCIM	HGYMSKMG	NP	FLTQ	W	Q	RRY	F	Y	L	P	N	R	L	E	W	R	G	E	G	E	A	P	Q	S	L	600																	
	NOV9	LTME	EIQ	S	VEET	Q	I	K	E	R	K	L	L	K	I	R	G	G	K	Q	F	I	L	Q	C	S	D	P	E	L	V	Q	W	K	E	L	R	D	A	Y	R	E	A	Q	L	V	Q	R	639
30	A53791	LTME	EIQ	S	VEET	Q	I	K	E	R	K	L	L	K	I	R	G	G	K	Q	F	I	L	Q	C	S	D	P	E	L	V	Q	W	K	E	L	R	D	A	Y	R	E	A	Q	L	V	Q	R	660
	P25098	LTME	EIQ	S	VEET	Q	I	K	E	R	K	L	L	K	I	R	G	G	K	Q	F	I	L	Q	C	S	D	P	E	L	V	Q	W	K	E	L	R	D	A	Y	R	E	A	Q	L	V	Q	R	660
	Q99LL8	LTME	EIQ	S	VEET	Q	I	K	E	R	K	L	L	K	I	R	G	G	K	Q	F	I	L	Q	C	S	D	P	E	L	V	Q	W	K	E	L	R	D	A	Y	R	E	A	Q	L	V	Q	R	658
	P26817	LTME	EIQ	S	VEET	Q	I	K	E	R	K	L	L	K	I	R	G	G	K	Q	F	I	L	Q	C	S	D	P	E	L	V	Q	W	K	E	L	R	D	A	Y	R	E	A	Q	L	V	Q	R	660
35	NOV9	VPKMK	NK	PRSP	V	V	E	L	S	K	V	P	L	V	Q	R	G	S	A	N	G	L	668																										
	A53791	VPKMK	NK	PRSP	V	V	E	L	S	K	V	P	L	V	Q	R	G	S	A	N	G	L	689																										
	P25098	VPKMK	NK	PRSP	V	V	E	L	S	K	V	P	L	V	Q	R	G	S	A	N	G	L	689																										
	Q99LL8	VPKMK	NK	PRSP	V	V	E	L	S	K	V	P	L	V	Q	R	G	S	A	N	G	L	687																										
	P26817	VPKMK	NK	PRSP	V	V	E	L	S	K	V	P	L	V	Q	R	G	S	A	N	G	L	689																										

Tables 9E-9L list the domain descriptions from DOMAIN analysis results against NOV9. This indicates that the NOV9 sequence has properties similar to those of other proteins known to contain this domain.

Table 9E. Domain Analysis of NOV9

gnl|Smart|smart00220, S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. (SEQ ID NO:98)
CD-Length = 256 residues, 100.0% aligned
Score = 237 bits (604), Expect = 2e-63

45

Query: 191 FSVHRIIGRGGFGEVYGCRAKADTGKMYAMKCLDKKRIKMQGETLALNERIMLSLVSTGD 250
+ + ++|+| |++| | |++|+|+|+++|+| | |+| + |
Sbjct: 1 YELLEVLGKGAFGKVYLARDKKTGKLVAIKVIKKEKLKKKKRER-ILREIKILKKL---D 56

50

Query: 251 CPFIVCMSYAFHTPDKLSFILDMNGGDLHYHLSQHGVSFSEADMRFYAAEIIILGLEHMHN 310
| | | + | | | |+++ | | | | + | | | + | | | |++ |++|+
Sbjct: 57 HPNIVKLYDVFEDDDKLYLVMEYCEGGDLFDLLKKRGRLESEARFYARQILSALEYLHS 116

55

Query: 311 RFVYRDLKPANILLDEHGHVRISDLGLACDFSKKKPHAS--VGTHGYMAPEVLQKGVA 368
+ +++| | | | | | | | | |++| | | | + | | | | | | | | | |

Sbjct: 117 QGIIHRDLKPENILLDSGDGHVKLADFGLAKQLDSGGTLLTTFVGTPEYMAPEVL-LGKGY 175

Query: 369 DSSADWFSLGCMLEKLLRGHSPFRQHKTKDK-HEIDRMTLTMAVELPDSFSPELRSLEG 427
 + | +||| +|++| | | | +

5 Sbjct: 176 GKAVDIWSLGVILYELLTGKPPFGDDQLLALFKKIGKPPPPFPPPEWKISPEAKDLIKK 235

Query: 428 LLQRDVNRRLGCLGRGAQEVKESPPF 453
 || +| +|| | +| | |||

10 Sbjct: 236 LLVKDPEKRL-----TAEAELEHPFF 256

Table 9F. Domain Analysis of NOV9

gnl|Pfam|pfam00069, pkinase, Protein kinase domain. (SEQ ID NO:99)
 CD-Length = 256 residues, 100.0% aligned
 Score = 221 bits (562), Expect = 1e-58

Query: 191 FSVHRIIGRGGEVYGCRAKADTGKMYAMKCLDKKRIKMKQGETLALNERIMLSLVSTGD 250
 + + +| | | +| + |||++ | + | + + | + | +| +|

15 Sbjct: 1 YELGEKLGSGAFGKVYKGGKHKDTGEIVAIIKKRSLSEKKRFL--REIQILRRLS--- 55

Query: 251 CPFIVCMSYAFHTPDKLSFILDLMNGGDLHYHLSQHGVF-SEADMRFYAAEIIIGLEHMH 309
 | || + | | | +| + ||| +| ++| + | + + | +| + |||++|

Sbjct: 56 HPNIVRLGVLFEEDDHLYLVMEYMEGGDLFDYLRNGLLLSEKEAKKIALQILRGLEYLH 115

20 Query: 310 NRFVVYRDLKPANILLDEHGHVRIIDLGLACDF---SKKKPHASVGTGYMAPEVLQKGV 366
 +| +| +||| | ||| +| +| +| | | | | | +|

Sbjct: 116 SRGIVHRDLKPENILLDENGTVKIDFGLARKLESSSYEKLTTFVGTPEYMAPEVL-EGR 174

25 Query: 367 AYDSSADWFSLGCMLEKLLRGHSPFRQHKTKDKHEIDRMTLTMAVELPDSFSPELRSLE 426
 | | | +||| +|++| | | | ++ + + + | + | | +| ++

Sbjct: 175 GYSSKVDVWSLGVILYELLTGKLPFGIDPLEELFRIKERERLRLPLPPNCSEELKDLIK 234

Query: 427 GLLQRDVNRRLGCLGRGAQEVKESPPF 453
 | +| +| | +| +| +|

30 Sbjct: 235 KCLNKDPEKRP-----TAKEILNHPWF 256

Table 9G. Domain Analysis of NOV9

gnl|Pfam|pfam00615, RGS, Regulator of G protein signaling domain. RGS family members are GTPase-activating proteins for heterotrimeric G-protein alpha-subunits. (SEQ ID NO:106)
 CD-Length = 119 residues, 100.0% aligned
 Score = 130 bits (326), Expect = 3e-31

Query: 54 TFEKIFSQKLGYYLFRDFCLNHLEEARPLVEFYEEIKKYEKLETEEERVARSRIFDSYI 113
 +|||+ | +| |||+| | +||+ +++||| | ++| ++|||+| +|

35 Sbjct: 1 SFEKLLKQPIGRLLFRELETFE--ENLEFWLAVEEYEKTEDPKRPPDKAREIYDEFI 58

Query: 114 MKELLACSHPFSSKATEHVQCHLGKKQVPPDLFQPYIEEICQNLRGDVFQKFIESDKFTR 173
 | | | +| | | | +| | | +||| | +| +||| |||

40 Sbjct: 59 SPEAPKPEVNLDSELREHTQDNL-LKAPTDFEQAQREIYDLMRGDSFPRFLESDFTR 117

Query: 174 FC 175
 |

Sbjct: 118 FL 119

Table 9H. Domain Analysis of NOV9

gnl|Smart|smart00219, TyrKc, Tyrosine kinase, catalytic domain;
Phosphotransferases. Tyrosine-specific kinase subfamily. (SEQ ID
NO:100)

CD-Length = 258 residues, 94.6% aligned
Score = 110 bits (275), Expect = 3e-25

	Query:	195	RIIGRGGFGEVYGCR---KADTGKMYAMKCLDKKRIKMKQGETLALNE-RIMLSLVSTGD	250
			+ + + + +	
5	Sbjct:	5	KKLGEGAFGEVYKGTCLKGGGVEVEVAVKTL--KEDASEQQIEEFLREARLMRKL----D	58
	Query:	251	CPFIVCMSYAFHTPDKLSFILDMNGGDLHYHLSQH--VFSEADMRFYAAEIIILGLEHM	308
			+ + +++ + ++ ++ + + + +++	
	Sbjct:	59	HPNIVKLLGVCTEEEEPIMIVMEYMEGGDLLDYLRKNRPKELSLSDLFSFALQIARGMEYL	118
10	Query:	309	HNRFVVYRDLKPANILLDEHGHVRISDLGLACDFSKKKPHASVGTGTHG----YMAPEVLQK	364
			++ + + + + + + + + +	
	Sbjct:	119	ESKNFVHRDLAARNCLVGENKTVKIADFLGLARDLYDDDYRKKKSPRLPIRWMAPESLKD	178
	Query:	365	GVAYDSSADWFSGLCMLFKLL-RGHSPFRQHKTKDKKHEIDRMTLTMAVELPDSFSPELRS	423
15			+ + + + +++ + ++ ++ + + + +	
	Sbjct:	179	GK-FTSKSDVWSFGVLLWEIFTLGESPY--PGMSNEEVLEYLKKGYRLPQPPNCPDEIYD	235
	Query:	424	LLEGLLQRDVNR	436
			+	
20	Sbjct:	236	LMQCWAEDPEDR	248

Table 9I. Domain Analysis of NOV9

gnl|Smart|smart00315, RGS, Regulator of G protein signalling domain;
RGS family members are GTPase-activating proteins for heterotrimeric
G-protein alpha-subunits. (SEQ ID NO:107)

CD-Length = 119 residues, 100.0% aligned
Score = 100 bits (248), Expect = 3e-22

	Query:	54	TFEKIFSQKLGYLFRDFCLNHLEEARPLVEFYEEIKKYEKLETEEERVARSR EIFDSYI	113
25	Sbjct:	1	+ + + + + + + +++++ +++++ + +	
			SLESLLRDPIGRLLFRE FLESEFSE--ENLEFWLAVEEFKKAED EERRSKAKEIYDKYL	58
	Query:	114	MKELLACSHPFKSATEHVQGH LGKKQVP PDLFQPYIEEICQNLRGDV FQKFIESDKFTR	173
30	Sbjct:	59	++ + ++ + + + + + +	
			SPNAPKE-VNLDS DLREEIEENLNKNEEPPDLFDEAQEEVYELLEKDSYPRFLES DYLLR	117
	Query:	174	FC 175	
	Sbjct:	118	FL 119	

Table 9J. Domain Analysis of NOV9

gnl|Smart|smart00233, PH, Pleckstrin homology domain.; Domain commonly found in eukaryotic signalling proteins. The domain family possesses multiple functions including the abilities to bind inositol phosphates, and various proteins. PH domains have been found to possess inserted domains (such as in PLC gamma, syntrophins) and to be inserted within other domains. Mutations in Brutons tyrosine kinase (Btk) within its PH domain cause X-linked agammaglobulinaemia (XLA) in patients. Point mutations cluster into the positively charged end of the molecule around the predicted binding site for phosphatidylinositol lipids. (SEQ ID NO:108)
 CD-Length = 104 residues, 95.2% aligned
 Score = 62.0 bits (149), Expect = 1e-10

Query: 539 IMHGYMSKMGNPFLTQWQRRYFYLFPPNRLEW-----RGEGEAPQSLLTMEEIQ---SVEE 590
 | |++ | + |++|| | | + + | + + + + +
 Sbjct: 2 IKEGWLLKKSSGGKSKWKRYFVLFNGVLLYYKSKKKSSSKPKGSIPLSGCTVREAPDS 61

Query: 591 TQIKERKCLLLKIRGGKQFILQCDSDPQLVQWKELRDA 629
 |++ | + | +|| +| + | + | + || |
 Sbjct: 62 DSDKKKNCFEIVTPDRKTLLQAESEERKEWVEALRKA 100

Table 9K. Domain Analysis of NOV9

gnl|Pfam|pfam00169, PH, PH domain. PH stands for pleckstrin homology. (SEQ ID NO:109)
 CD-Length = 100 residues, 97.0% aligned
 Score = 55.5 bits (132), Expect = 1e-08

Query: 539 IMHGYMSKMGNPFLTQWQRRYFYLFPPNRLEW---RGEGEAPQSLLTMEEIQSVEETQIKE 595
 + |++ | +|++|||+|| + | + + + | + + + + +
 Sbjct: 2 VKEGWLLKKSTVKKRKKRYFVLFNDVLIYYKDKKKSYPKGSIPLSGCSVEDVPDSEF 61

Query: 596 RKCLLLKIR---GGKQFILQCDSDPQLVQWKELRDA 629
 ++ ++| | + ||| +| + | | | ++ |
 Sbjct: 62 KRPNCQLRSRDGKETFILQAESEERQDWIKAIQSA 98

Table 9L. Domain Analysis of NOV9

gnl|Smart|smart00133, S_{TK}X, Extension to Ser/Thr-type protein kinases (SEQ ID NO:110)
 CD-Length = 63 residues, 87.3% aligned
 Score = 42.7 bits (99), Expect = 7e-05

Query: 454 RSLDWQMVFLOKYPPPLIPPRGEVNAADAFDIGSFDEEDTKGIKQVEAETVFDTINAETD 513
 | +|| + ++ || +| | +|| | | ++ | | +|+|
 Sbjct: 1 RGIDWDKLENKEIEPPFVPVKV-----SPTDTSNFDPEFT---EESPVLTPVDPPLSESD 52

Query: 514 RLE 516
 + |
 Sbjct: 53 QDE 55

Eukaryotic protein kinases are enzymes that belong to a very extensive family of proteins which share a conserved catalytic core common with both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. In the N-terminal extremity of the catalytic domain there is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. In the central part of the catalytic domain there is a conserved aspartic acid residue which is important for the catalytic activity of the enzyme.

The beta-adrenergic receptor kinase (beta ARK) catalyses the phosphorylation of the activated forms of the beta 2-adrenergic receptor (beta 2AR). The interaction between receptor and kinase is independent of second messengers and appears to involve a multipoint attachment of kinase and substrate with the specificity being restricted by both the primary amino acid sequence and conformation of the substrate. Kinetic, functional and sequence information reveals that rhodopsin kinase and beta ARK are closely related, suggesting they are members of a family of G-protein-coupled receptor kinases.

The beta-adrenergic signaling cascade is an important regulator of myocardial function. Significant alterations of this pathway are associated with several cardiovascular diseases, including congestive heart failure (CHF). CHF patients share several similar features, such as reduced cardiac contractility and neurohumoral activation to compensate the impaired cardiac function. In CHF patients, the cardiac renin-angiotensin (RA) system, receptors, GTP-binding proteins, and their effector molecules are inevitably exposed to chronically elevated neurohumoral stimulation. A widely recognized concept is that a chronic increase in such stimulation can desensitize target cell receptors and the post-receptor signal transducing pathway. Included in these alterations is increased activity and expression of G protein-coupled receptor kinases (GRKs), such as the beta-adrenergic receptor kinase (beta ARK1), which phosphorylate and desensitize beta-adrenergic receptors (beta ARs). A body of evidence is accumulating that suggests that GRKs, in particular beta ARK1, are critical determinants of cardiac function under normal conditions and in disease states. Transgenic mice with myocardial-targeted alterations of GRK activity have shown profound changes in the in vivo functional performance of the heart. Included in these studies is the compelling finding that inhibition of beta ARK1 activity or expression significantly enhances cardiac function and potentiates beta AR signaling in failing cardiomyocytes. An uncoupling of beta2-adrenoceptors has been attributed to an increased activity and gene expression of beta-adrenergic receptor kinase in failing myocardium, leading to phosphorylation and uncoupling

of receptors. The important physiological function of GRK2 as a modulator of the efficacy of GPCR signal transduction systems is exemplified by its relevance in cardiovascular physiopathology as well as by its emerging role in the regulation of chemokine receptors.

The disclosed NOV9 nucleic acid of the invention encoding a Beta-adrenergic receptor kinase-like protein includes the nucleic acid whose sequence is provided in Table 9A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 9A while still encoding a protein that maintains its Beta-adrenergic receptor kinase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 2 percent of the bases may be so changed.

The disclosed NOV9 protein of the invention includes the Beta-adrenergic receptor kinase-like protein whose sequence is provided in Table 9B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 9B while still encoding a protein that maintains its Beta-adrenergic receptor kinase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 1 percent of the residues may be so changed.

The protein similarity information, expression pattern, and map location for the beta-adrenergic receptor kinase-like protein and the NOV9 proteins disclosed herein suggest that this beta-adrenergic receptor kinase may have important structural and/or physiological functions characteristic of the Ser/Thr protein kinases family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene

delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon.

The NOV9 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from heart failure, hypertension, secondary pathologies caused by heart failure and hypertension, and other diseases, disorders and conditions of the like. Additionally, the compositions of the present invention may have efficacy for treatment of patients suffering from conditions associated with the role of GRK2 in brain and in the regulation of chemokine receptors.. The NOV9 nucleic acid, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV9 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV9 protein have multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, contemplated NOV9 epitope is from about amino acids 40 to 70. In another embodiment, the contemplated NOV9 epitope is from about amino acids 80 to 110. In further embodiments, the contemplated NOV9 epitope is from about amino acids 120 to 200, from about amino acids 220 to 245, from about amino acids 250 to 280, or from about amino acids 290 to 340. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV10

A disclosed NOV10 nucleic acid of 3003 nucleotides (also referred to as AC058790_da25) encoding an alpha-mannosidase-like protein is shown in Table 10A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 57-59 and ending with a TAA codon at nucleotides 2946-2948. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 10A. The start and stop codons are in bold letters. Single nucleotide polymorphism data is included in Example 3.

Table 10A. NOV10 nucleotide sequence (SEQ ID NO:25).

```

GGTATCATACTCCAGCAAGCGCACATCATCAGTGACGTGATCAGGATGCATCGTCATGGCGGCAGCGCCGTTCTTGAAG
CACTGGCGCACCACCTTTTGAAGCGGGTGGAGAAGTTTCGTGTCCCGGATCTACTTCACCGACTGTAACCTCCGCGGCAGGCT
TTTTGGGGCCAGCTGCCCTGTGGCTGTGCTCTCCAGCTTCTTGACGCCGGAGAGACTTCCCTACCAGGAGGCAGTCCAGC
GGGACTTCCGCCCCCGCGCAGGTCCGGCGACAGCTTCGGACCCACATGGTGGACCTGCTGGTTCGGGTGGAGCTGACCATC
CCAGAGGCATGGGTGGGCCAGGAAGTTCACCTTTGCTGGGAAAGTGATGGAGAAGGTCTGGTGTGGCGTGATGGAGAACC
TGTCCAGGGTTTAAACAAAGAGGGTGAGAAGACCAGCTATGTCTGACTGACAGGCTGGGGGAAAGAGACCCCCGAGCC
TCACCTCTATGTGGAAGTAGCCTGCAATGGGCTCCTGGGGGCCGGGAAGGGAAGCATGATTGCAGCCCCGTGACCTGAG
AAGATGTTCCAGCTGAGCCGGCTGAGCTAGCTGTGTTCCACCGGATGTCCACATGCTCCTGGTGGATCTGGAGCTGCT
GCTGGGCATAGCCAAGGCGCAGCAGCTGGAATGGGTGAAGAGCCGTACCCTGGCCTGTACTCCCGCATCCAGGAGTTTG
CGTGCCGTGGGCAGTTTGTGCTGTGGGGGACCTGGGTGGAGATGGATGGGAACCTGCCAGTGGAGAGGCCATGGTG
AGGCAGTTTTTGCAGGGCCAGAAGCTTCTTCTGCAGGAGTTTGGGAAGATGTGCTCTGAGTTCTGGCTGCCGGACACCTT
TGGCTACTCAGCACAGCTCCCCAGATCATGCACGGCTGTGGCATCAGGCGCTTCTCACCCAGAAATTGAGGAAATT
TGGTGAAGCTCCTTCCACACCATACATTTTTCTGGGAGGGCCCTGGATGGCTCCCGTGTACTGGTCCACTTCCACCTGGC
GACTCCTATGGGATGCAGGGCAGCGTGGAGGAGTGTGAAGACCGTGGCCAAACACCGGACAAGGGCGGGCCAAACA
CAGTGCCCTTCTCTTTGGCTTTGGGGATGGGGGTGGTGGCCCCACCCAGACCATGCTGGACCGCTGAAGCGCCTGAGCA
ATACGGATGGGCTGCCAGGCTGACAGTATCTTCTCAAGACAGCTCTTCTCAGCACTGGAGAGTGAAGTCAAGAGCAGTG
TGCAGTGGGTGGGGAGCTCTTCTTGGAGCTGCACAAATGGCACATACACCCATGCCAGATCAAGAAGGGGAACCG
GGAATGTGAGCGGATCTGCACGACGTGGAGCTGCTCAGTAGCCTGGCCCTGGCCCGCAGTGGCCAGTTCTATACCCAG
CAGCCAGCTGCAGCACCTCTGGAGGCTCCTTCTTCTGAACAGTTCCATGATGTGGTGAAGCTGCATCCAGATG
GTGGCAGAGGAAGCCATGTGCCATTATGAAGACATCCGTTCCCATGGCAATACACTGCTCAGCGCTGCAGCCGAGCCCT
GTGTGCTGGGAGCCAGTCTGTAGGCGCTCCTCATCGTCAACACACTGCCCTGGAAGCGGATCGAAGTGATGGCCCTGC
CCAAACCGGGCGGGGCCACAGCCTAGCCCTGGTGACAGTGGCCAGCATGGGCTATGCTCCTGTCTCCCCACCTCA
CTGCAGCCCCTGTGCCCCAGCAGCCTGTGTTCTGTAGTCAAGAGACTGATGGTCCGTGACTCTGGACAATGGCATCAT
CCGAGTGAAGCTGGACCAACTGGTGCCTGACGTCTTGGTCTGGTGGCCTCTGGCAGGGAGGCCATTGCTGAGGGCG
CCGTGGGGAACAGTTTGTGCTATTGATGATGTCCCTTGTACTGGGATGCATGGGACGTGATGGAATACCACTGGAG
ACACGGAAGCCTGTGCTGGGCCAGGCAGGACCTGGCAGTGGGCACCGAGGGCGGCCCTGCGGGCAGCGCCTGGTTCTT
GCTACAGATCAGCCCCAACAGTCCGGCTTAGCCAGGAGTTGTGCTGGACGTTGGCTGCCCTATGTCCGCTTCCACACCG
AGGTACACTGGCATGAGGCCACAAGTTCTGAAGGTGGAGTTCCTGCTCGCTGCGGAGTTCCAGGCCACCTATGAG
ATCCAGTTTGGGCACCTGCAGCGACCTACCCACTACAATACCTCTTGGACTGGGCTCGATTGAGGTGTGGGCCCATCG
CTGGATGGATCTGTGCAACACCGGCTTGGGCTGGCCCTGAGGCTCAACGACTGCAAGTATGGCGCGTCAGTGGAGGAGCA
TCCTCAGCCTCTCGCTTGTGGGGCGCCTAAAGCCCCGACGCTACTGCTGACACGGGGCGCCACGAGTTACCTATGCA
CTGATGCCGCAAGGGCTCTTCCAGGATGCTGGCGTTATCCAAGCTGCCTACAGCCTAAACTTCCCCCTGTTGGCTCT
GCCAGCCCCAGCCAGCGCCGCCACCTCCTGGAGTGCCTTTCCGTGTCTTACCCGCGGTCGTATTGGAGACCGTCA
AGCAGGCGGAGAGCAGCCCCCAGCGCGCTCGCTGGTCTGAGGCTGTATGAGGCCACGGCAGCCACGTGCTGG
CTGCACTTGTGCTGCCGTTTCAAGAGGCCATCCTCTGCGATCTTGGAGCGACCGAGCCCTGCTGGCCACTTGACTTC
GGGACAACCGCTGAAGCTCACCTTTCTCCCTTCCAAGTGTCTCCCTGTTGCTCGTGCTTCAAGCCTCCGCCACACTGA
GTCCTTGGGGCTGGGGTTTTGTTGTAGAGGCTCTGGGACTCCTAATTTCTGCTTCCCAGCCTAAAGCAGGGATCAG
TCTTTTCTTGTGGAATAAATCCTTGGATCGGAAAAA

```

In a search of public sequence databases, the NOV10 nucleic acid sequence, located on chromosome 15 has 2371 of 2390 bases (99%) identical to a alpha-mannosidase mRNA from *Homo sapiens*, (GENBANK-ID: AF044414| acc: AF044414.2) (E = 0.0). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV10 polypeptide (SEQ ID NO:26) encoded by SEQ ID NO:25 has 963 amino acid residues and is presented in Table 10B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV10 does not have a signal peptide and is likely to be localized in the peroxisome (microbody) with a certainty of 0.7480. In other embodiments, NOV10 is also likely to be localized to the mitochondrial membrane space with a certainty of 0.4539, to the mitochondrial intermembrane space with a certainty of 0.4027, or to the lysosome (lumen) with a certainty of 0.2317.

Table 10B. Encoded NOV10 protein sequence (SEQ ID NO:26).

```

MAAAPFLKHWRTTFFERVEKFVSPITYFTDCNLRGRLEFGASCFAVLSSFLTPERLPYQEAQVQDRFPAQVGD
FGPTWWTWQFVERVELTPEAWVGQEVHLCWESDGEGLVWRDGEFVQGLTKEGEKTSYVLTDRLGERDPRSLTL
YVEVACNGLLGAGKGSMAAPDPEKMFQLSRAELAVFHRDVHMLLVLDLELLGIAKAQQLEWVKSRYPGLYS
RIQEFACRGQFVPGGTWVEMDGNLPSGEAMVRQFLQGNFFLQEFKMCSEFWLPDTFGYSAQLPQIMHGC
GIRRFLLTQKLSWNLVNSFFPHHTFFWEGLDGSRVLVHFPPGDSYGMQGSVEEVLTVANNRDGRANHSALFLF
GFGDGGGGPTQTMLDRLKRLSNTDGLPRVQLSSPRQLFSALESQDSEQLCTWVGELFLELHNGTYTTTHAIKK
GNRECEILHDVELLSLALARSQFLYPAAQLQLWRLLLLNQFHDVVTGSCIQMVAAEAMCHYEDIRSHG
NTLLSAAAAALCAGEPGPEGLLIVNTLPWKRIEVMALPKPGGAHSLALVTVPSMGYAPVPPPTSLQPLLPQQ
PVFVVQETDGSVTLNNGIIRVKLDPTGRLTSLVLVASGREATAEGAVGNQFVLFDDVPLYWDAWDVMDYHLE
TRKPVLGQAGTLAVGTEGGLRGSAWFLQISPNRSLSQEVVLDVGCPCYVRFHTEVHWHEAHKFLKVEFFARV
RSSQATYEQFQHLQRPHTYNTSWDWARFEVWAHRWMDLSEHGFGALLNDCKYGASVRGSILSLSLRAPK
APDATADTGRHEFTYALMPHKGSFQDAGVIAAYSLNFLLALPAPSPAPATSWSAFVSVPVAVVLETVKQA
ESSPQRRSLVLRLYEAHGSVHDCWLHLSLPVQEAILCDLLERPDAGHLTSGQPPEAHLFSLPASVAVPARAS
ASATLSPWGWGFVCCRLLWGLLISASPA

```

A search of sequence databases reveals that the NOV10 amino acid sequence has 764 of 771 amino acid residues (99%) identical to, and 767 of 771 amino acid residues (99%) similar to, the 1062 amino acid residue alpha-mannosidase protein from *Homo sapiens* (Q9UL64) (E = 0.0). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV10 was derived from a pool of the following tissues: Adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain - whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus, Bone, Cervix, Chorionic Villus, Colon, Liver, Lung, Lymph node, Lymphoid tissue, Ovary, Peripheral Blood, Skin, Stomach, Tonsils, Whole Organism. Thus, it is expressed in at least some of the above tissues. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Genomic Clone sources, Literature sources, and/or RACE sources.

The disclosed NOV10 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 10C.

Table 10C. BLAST results for NOV10

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
ptnr: SPTREMBL-ACC: Q9UL64	ALPHA MANNOSEDASE 6A8B - Homo sapiens	1062	763/771 (99%)	767/771 (99%)	0.0

ptnr:SPTREMBL- ACC:Q9NTJ4	HYPOTHETICAL 115.8 KDA PROTEIN - Homo sapiens	1040	715/722 (99%)	718/722 (99%)	0.0
ptnr:TREMBLNEW- ACC:AAH16253	SIMILAR TO MANNOSIDASE, ALPHA, CLASS 2C, MEMBER 1	1039	635/730 (89%),	692/730 (94%)	0.0
ptnr:SWISSPROT- ACC:P21139	Alpha-mannosidase (EC 3.2.1.24)	1040	625/731 (85%)	661/731 (90%)	0.0
ptnr:SPTREMBL- ACC:Q13358	ALPHA-MANNOSIDASE - Homo sapiens	425	425/425 (100%)	425/425 (100%)	0.0

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 10D. In the ClustalW alignment of the NOV10 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 10D. ClustalW Analysis of NOV10

- 1) NOV10 (SEQ ID NO:26)
- 2) ptnr: ALPHA MANNOSIDASE 6A8B - Homo sapiens (SEQ ID NO:73)
- 3) ptnr: HYPOTHETICAL 115.8 KDA PROTEIN - Homo sapiens (SEQ ID NO:74)
- 4) ptnr: SIMILAR TO MANNOSIDASE, ALPHA, CLASS 2C, MEMBER 1 (SEQ ID NO:75)
- 5) ptnr: Alpha-mannosidase (EC 3.2.1.24) (SEQ ID NO:76)

NOV10	MAAAPFLKHHRTTFFERVEKVSPIYFTDNLGRRLFGASCPVAVLSSFLTIERLPYQEAQVRI-----FRPAQVQ	70
Q9UL64	MAAAPFLKHHRTTFFERVEKVSPIYFTDNLGRRLFGASCPVAVLSSFLTIERLPYQEAQVRI-----FRPAQVQ	70
Q9NTJ4	MAAAPFLKHHRTTFFERVEKVSPIYFTDNLGRRLFGASCPVAVLSSFLTIERLPYQEAQVRI-----FRPAQVQ	70
AAH16253	MAAAPFLKHHRTTFFERVEKVSPIYFTDNLGRRLFGASCPVAVLSSFLTIERLPYQEAQVRI-----FRPAQVQ	69
P21139	MAAAPFLKHHRTTFFERVEKVSPIYFTDNLGRRLFGASCPVAVLSSFLTIERLPYQEAQVRI-----FRPAQVQ	79
NOV10	ISFGPTTWTCWFRVELTIPEAVWGQEVHLCWEISDGEGLVWRIGEPVQGLTH-----EGEKTSYVLIDRLGERDPF	140
Q9UL64	ISFGPTTWTCWFRVELTIPEAVWGQEVHLCWEISDGEGLVWRIGEPVQGLTH-----EGEKTSYVLIDRLGERDPF	140
Q9NTJ4	ISFGPTTWTCWFRVELTIPEAVWGQEVHLCWEISDGEGLVWRIGEPVQGLTH-----EGEKTSYVLIDRLGERDPF	140
AAH16253	ISFGPTTWTCWFRVELTIPEAVWGQEVHLCWEISDGEGLVWRIGEPVQGLTH-----EGEKTSYVLIDRLGERDPF	139
P21139	ISFGPTTWTCWFRVELTIPEAVWGQEVHLCWEISDGEGLVWRIGEPVQGLTH-----EGEKTSYVLIDRLGERDPF	159
NOV10	SLTLYVEVACNGLLGAGKGSMTAAPDPEKHHQLSRAELAVF-----HRDVHMLLVLLELLLGIAH	200
Q9UL64	SLTLYVEVACNGLLGAGKGSMTAAPDPEKHHQLSRAELAVF-----HRDVHMLLVLLELLLGIAH	211
Q9NTJ4	SLTLYVEVACNGLLGAGKGSMTAAPDPEKHHQLSRAELAVF-----HRDVHMLLVLLELLLGIAH	211
AAH16253	SLTLYVEVACNGLLGAGKGSMTAAPDPEKHHQLSRAELAVF-----HRDVHMLLVLLELLLGIAH	210
P21139	SLTLYVEVACNGLLGAGKGSMTAAPDPEKHHQLSRAELAVF-----HRDVHMLLVLLELLLGIAH	229
NOV10	ALYTANQMNVNLPDPAQPETFVYQALASRFFGQHGGSQHTIHATGHCHIDTAWLWPFKETVRKCARSWVATQIMERNP	290
Q9UL64	ALYTANQMNVNLPDPAQPETFVYQALASRFFGQHGGSQHTIHATGHCHIDTAWLWPFKETVRKCARSWVATQIMERNP	291
Q9NTJ4	ALYTANQMNVNLPDPAQPETFVYQALASRFFGQHGGSQHTIHATGHCHIDTAWLWPFKETVRKCARSWVATQIMERNP	291
AAH16253	ALYTANQMNVNLPDPAQPETFVYQALASRFFGQHGGSQHTIHATGHCHIDTAWLWPFKETVRKCARSWVATQIMERNP	290
P21139	ALYTANQMNVNLPDPAQPETFVYQALASRFFGQHGGSQHTIHATGHCHIDTAWLWPFKETVRKCARSWVATQIMERNP	229
NOV10	-----AQQLLEWVKSRYPGLYSRIOEFACRGQFVPGGTVVEMDNLPSGEAMVRQFLQGNFFLQEFGRKMCSEFWLI	272
Q9UL64	-----AQQLLEWVKSRYPGLYSRIOEFACRGQFVPGGTVVEMDNLPSGEAMVRQFLQGNFFLQEFGRKMCSEFWLI	371
Q9NTJ4	-----AQQLLEWVKSRYPGLYSRIOEFACRGQFVPGGTVVEMDNLPSGEAMVRQFLQGNFFLQEFGRKMCSEFWLI	371
AAH16253	-----AQQLLEWVKSRYPGLYSRIOEFACRGQFVPGGTVVEMDNLPSGEAMVRQFLQGNFFLQEFGRKMCSEFWLI	370
P21139	-----AQQLLEWVKSRYPGLYSRIOEFACRGQFVPGGTVVEMDNLPSGEAMVRQFLQGNFFLQEFGRKMCSEFWLI	281
NOV10	DTFGYSAQLPQIMHGGGIRRFITQKLSNVLNVSFPHTTFEWEGLDGSRLVHFPFGDSYGMQGSVEEVLTQVNNKDRGI	352
Q9UL64	DTFGYSAQLPQIMHGGGIRRFITQKLSNVLNVSFPHTTFEWEGLDGSRLVHFPFGDSYGMQGSVEEVLTQVNNKDRGI	451
Q9NTJ4	DTFGYSAQLPQIMHGGGIRRFITQKLSNVLNVSFPHTTFEWEGLDGSRLVHFPFGDSYGMQGSVEEVLTQVNNKDRGI	451
AAH16253	DTFGYSAQLPQIMHGGGIRRFITQKLSNVLNVSFPHTTFEWEGLDGSRLVHFPFGDSYGMQGSVEEVLTQVNNKDRGI	450
P21139	DTFGYSAQLPQIMHGGGIRRFITQKLSNVLNVSFPHTTFEWEGLDGSRLVHFPFGDSYGMQGSVEEVLTQVNNKDRGI	314
NOV10	-----GESQHTIATGHCHIDTAWL-----P-----FKITVRKCTI	432

Table 10E lists the domain description from DOMAIN analysis results against

Table 10E. Domain Analysis of NOV10

118

			laggskiegasyAiKleklyeqleelRralaIfQHDAiTGTakqhVv<-*	
		+++++	+ + + + + + + + + + +	
AC058790_d	454	RS-AQFLYP-----A-----	QLQHLWRLLLLNQFHDVVTGSCIQMV	490

Glycosyl hydrolases are key enzymes of carbohydrate metabolism. Lysosomal alpha-mannosidase is necessary for the catabolism of N-linked carbohydrates released during glycoprotein turnover. The enzyme catalyzes the hydrolysis of terminal, non-reducing alpha-D-mannose residues in alpha-D-mannosides, and can cleave all known types of alpha-mannosidic linkages. While alpha-mannosidases were classified as enzymes that process newly formed N-glycans or degrade mature glycoproteins, two endoplasmic reticulum (ER) alpha-mannosidases with previously assigned processing roles, have important catabolic activities. The ER/cytosolic mannosidase may be involved in the degradation of dolichol intermediates that are not needed for protein glycosylation, whereas the soluble form of Man9-mannosidase is responsible for the degradation of glycans on defective or malformed proteins that are specifically retained and broken down in the ER. The degradation of oligosaccharides derived from dolichol intermediates by ER/cytosolic mannosidase explains why cats and cattle with alpha-mannosidosis store and excrete some unexpected oligosaccharides containing only one GlcNAc residue. Similarly, the action of ER/cytosolic mannosidase, followed by the action of the recently described human lysosomal alpha(1 → 6)-mannosidase, together explain why alpha-mannosidosis patients store and excrete large amounts of oligosaccharides that resemble biosynthetic intermediates, rather than partially degraded glycans. The relative contributions of the lysosomal and extra-lysosomal catabolic pathways can be derived by comparing the ratio of trisaccharide Man beta (1 → 4)GlcNAc beta (1 → 4)GlcNAc to disaccharide Man beta (1 → 4)GlcNAc accumulated in tissues from goats with beta-mannosidosis. A similar determination in human beta-mannosidosis patients is not possible because the same intermediate, Man beta (1 → 4)-GlcNAc is a product of both pathways. Based on inhibitor studies with pyranose and furanose analogues, alpha-mannosidases may be divided into two groups. Those in Class 1 are (1 → 2)-specific enzymes like Golgi mannosidase I, whereas those in Class 2, like lysosomal alpha-mannosidase, can hydrolyse (1 → 2), (1 → 3) and (1 → 6) linkages. A similar classification has been derived from protein sequence homologies. It is possible to speculate about their probable evolution from two primordial genes. The first would have been a Class 1 ER enzyme involved in the degradation of glycans on incompletely assembled or malformed glycoproteins. The second would have been a Class 2 lysosomal enzyme responsible for turnover. Later, other alpha-mannosidases, with new processing or catabolic functions, would have developed from these, by loss or gain

of critical insertion or retention sequences, to yield the full complement of alpha-mannosidases known today (Glycobiology 1994 Oct;4(5):551-66). Defects in the lysosomal alpha-mannosidase gene cause lysosomal alpha-mannosidosis (AM), a lysosomal storage disease characterized by the accumulation of unbranched oligo-saccharide chains. Depending on the clinical findings at the age of onset, a severe infantile (type I) and a mild juvenile (type II) form of alpha-mannosidosis are recognized. Furthermore, variability in clinical expression of the disease is seen within each type. Some of the disease features are: susceptibility to infection, vomiting, coarse features, macroglossia, flat nose, large clumsy ears, widely spaced teeth, large head, big hands and feet, tall stature, slight hepatosplenomegaly, muscular hypotonia, lumbar gibbus, radiographic skeletal abnormalities, dilated cerebral ventricles, lenticular opacities, hypogammaglobulinemia, 'storage cells' in the bone marrow, and vacuolated lymphocytes in the bone marrow and blood.

The disclosed NOV10 nucleic acid of the invention encoding a Alpha-mannosidase-like protein includes the nucleic acid whose sequence is provided in Table 10A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 10A while still encoding a protein that maintains its Alpha-mannosidase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 2 percent of the bases may be so changed.

The disclosed NOV10 protein of the invention includes the Alpha-mannosidase-like protein whose sequence is provided in Table 10B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 10B while still encoding a protein that maintains its Alpha-mannosidase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 1 percent of the residues may be so changed.

The protein similarity information, expression pattern, and map location for the alpha-mannosidase-like protein and the NOV10 protein disclosed herein suggest that this alpha-mannosidase-like protein may have important structural and/or physiological functions characteristic of the mannosidase protein family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These applications include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon.

The NOV10 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from alpha-mannosidosis, beta-mannosidosis, other storage disorders, peroxisomal disorders such as zellweger syndrome, infantile refsum disease, rhizomelic chondrodysplasia (chondrodysplasia punctata, rhizomelic), and hyperpipecolic acidemia and other diseases, disorders and conditions of the like. Since mannosidoses are found not only in humans, but also in animals, the nucleic acids and proteins of the this invention may be useful in treating animals with mannosidoses or other storage diseases, and other diseases, disorders and conditions of the like. Additionally, the compositions of the present invention may have efficacy for treatment of patients suffering from conditions associated with the role of GRK2 in brain and in the regulation of chemokine receptors.. The NOV10 nucleic acid, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV10 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV10 protein have multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, contemplated NOV10 epitope is from about amino acids 5 to 20. In another embodiment, the

contemplated NOV10 epitope is from about amino acids 40 to 80. In further embodiments, the contemplated NOV10 epitope is from about amino acids 110 to 180, from about amino acids 200 to 230, from about amino acids 300 to 370, from about amino acids 375 to 450, from about amino acids 650 to 680, from about amino acids 690 to 770, from about amino acids 790 to 820, from about amino acids 850 to 880, or from about amino acids 900 to 920. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV11

NOV11 includes three novel C1q-related factor-like proteins disclosed below. The disclosed sequences have been named NOV11a, NOV11b, and NOV11c. Single nucleotide polymorphism data is discussed below in Example 4.

NOV11a

A disclosed NOV11a nucleic acid of 805 nucleotides (also referred to as GM57107065_da1) encoding an C1q-related factor-like protein is shown in Table 11A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 83-85 and ending with a TGA codon at nucleotides 797-799. Putative untranslated regions are upstream from the initiation codon and downstream from the termination codon.

Table 11A. NOV11a nucleotide sequence (SEQ ID NO:27).

```
GAGTGAGGAAGATTGCTGGCCCTGGCAGCGTCGCGGCTGAGCCGCCGCAAGAGGGTGGCGGGCGCGCCGTCGGAGTGG
CCATGGTGCTGCTGCTGCTGGTGGCCATCCCGCTGCTGGTGACAGCTCCCGCGGGCCAGCGCACTACGAGATGCTGGGT
CGCTGCCGCATGGTGTGCGACCCGCATGGGCCCGTGGCCCTGGTCCGGACGGCGCGCCTGCTTCCGTGCCCCCTTCCC
GCCAGGCGCCAAGGGAGAGGTGGGCCGGTGCGGGAAAGCAGGCCCTGAGGGGGCCCCCTGGACCACCAGGTCCAAGAGGGC
CCCCAGGAGAACCCGGCAGGCCAGGCCCCCCGGGCCCTCCCGGTCCAGGTCCGGGCGGGGTGGCGCCCGCTGCCGGCTAC
GTGCCTCGCATTTGCTTTTACGCGGGCCTGCGGGGCCCCACGAGGGTTACGAGGTGCTGCGCTTCGACGACGTGGTGAC
CAACGTGGGCAACGCCTACGAGGCAGCCAGCGGCAAGTTTACTTGCCCCATGCCAGGCGTCTACTTCTTCGCTTACCACG
TGCTCATGCGCGCGCGGACGGCACCAGCATGTGGGCCGACCTCATGAAGAACGGACAGGTCCGGGCCAGCGCCATTGCT
CAGGACGCGGACCAGAACTACGACTACGCCAGCAACAGCGTCATTCTGCACCTGGACGTGGGCGACGAGGTCTTCATCAA
GCTGGACGGCGGAAAGTGCACGGCGGCAACACCAACAAGTACAGCACCTTCTCCGGCTTCATCATCTACCCCGACTGAG
CCGGC
```

In a search of public sequence databases, the NOV11a nucleic acid, located on chromosome 12, has 565 of 787 bases (71%) identical to a C1q-related factor mRNA from *Homo sapiens*, (GENBANK-ID: AF095154) ($E = 9.9e^{-68}$). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV11a polypeptide (SEQ ID NO:28) encoded by SEQ ID NO:27 has 238 amino acid residues and is presented in Table 11B using the one-letter amino acid code.

Signal P, Psort and/or Hydropathy results predict that NOV11a has a signal peptide and is likely to be localized extracellularly with a certainty of 0.5374. In other embodiments, NOV11a is also likely to be localized to the microbody (peroxisome) with a certainty of 0.1111, to the endoplasmic reticulum (membrane) with a certainty of 0.1000, and to the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV11a is between positions 15 and 16: VHS-SR.

Table 11B. Encoded NOV11a protein sequence (SEQ ID NO:28).

MVLLLLVAIPLLVHSSRGPAPHYEMLGRCRMVCDPHGPRGPGPDGAPASVPPFPPGAKGEVGRGKAGLRGPP GPPGPRGPPGEPGRPGPPGPPGPGGVAPAAGYVPRIAFYAGLRRPHEGYEVLRFDDVVTNVGNAYEAASG KFTCPMPGVYFFAYHVMRGGDGTSMWADLMKNGQVRASAIQDADQNYDYASNSVILHLDVGDEVFIKLDG GKVHGGNTNKYSTFSGFLIYPD
--

A search of sequence databases reveals that the NOV11a amino acid sequence has 184 of 258 amino acid residues (71%) identical to, and 198 of 258 amino acid residues (76%) similar to, the 258 amino acid residue C1q-related factor precursor protein from *Homo sapiens* (075973) ($E = 9.1 \times 10^{-91}$). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV11a is specifically expressed in the following tissues: brain, heart, testis, kidney, thyroid, prostate, fetal kidney, fetal skeletal. It shows increased expression in cancer cell lines derived from the following tissue: colon, kidney, ovary, skin, brain. It is highly upregulated in IFN-gamma treated endothelial cells. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources and Taqman results.

NOV11b

A disclosed NOV11b nucleic acid of 805 nucleotides (also referred to as CG54503-02) encoding a novel C1q-related factor-like protein is shown in Table 11C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 83-85 and ending with a TGA codon at nucleotides 797-799. Putative untranslated regions are underlined and are found upstream from the initiation codon and downstream from the termination codon.

Table 11C. NOV11b nucleotide sequence (SEQ ID NO:29).

GAGTGAGGAAGATTTGCTGGCCCTGGCAGCGTCGCGGGCTGAGCCGCCCAAGAGGGTGGCGGGCGGGCCGTCGGAGTGG
 CCATGGTGCTGCTGCTGCTGGTGGCCATCCCGCTGCTGGTGCACAGTCCCGCGGGCCAGCGCACTACGAGATGCTGGGT
 CGCTGCCGCATGGTGTGCGACCCGCATGGGCCCCGTGGCCCTGGTCCGGACGGCGCGCCTGCTTCCGTGCCCCCTTCCC
 GCCAGGCGCCAAGGGAGAGGTGGGCCGCGCGGGAAGCAGGCTGCGGGGGCCCCCTGGACCACCAGGTCCAAGAGGGC
 CCCAGGAGAACC CGGCAGGCCAGGCCCCCCCGGGCCCTCCCGTCCAGGTCCGGCGGGGTGGCGCCCGCTGCCGGCTAC
 GTGCCCTCGCATTGCTTTCTACGCGGGCTGCGCGGGCCCCACGAGGGTTACGAGGTGCTGCGCTTCGACGACGTGGTGAC
 CAACGTGGGCAACGCCTACGAGGCAGCCAGCGCAAGTTACTTGGCCCATGCCAGGCGTCTACTTCTTCGCTTACCACG
 TGCTCATGCGCGGGCGGACGGCACCAGCATGTGGGCCGACCTCATGAAGAACGGACAGGTCCGGGCCAGCGCCATTGCT
 CAGGACGCGGACCAGAACTACGACTACGCCAGCAACAGCGTCATTCTGCACCTGGACGTGGGCGACGAGGTCTTCATCAA
 GCTGGACGGCGGGAAGTGACGGCGGCAACCAACAAGTACAGCACCTTCTCCGGCTTCATCATCTACCCGACTGAG
 CCGGC

In a search of public sequence databases, the NOV11a nucleic acid, located on
 chromosome 17q21, has 565 of 787 bases (71%) identical to a C1q-related factor mRNA from
Homo sapiens, (GENBANK-ID: AF095154) ($E = 1.9e^{-68}$). Public nucleotide databases include
 all GenBank databases and the GeneSeq patent database.

The disclosed NOV11b polypeptide (SEQ ID NO:30) encoded by SEQ ID NO:29 has
 238 amino acid residues and is presented in Table 11D using the one-letter amino acid code.
 The SignalP, Psort and/or Hydropathy profile for NOV11b predict that this sequence has a
 signal peptide and is likely to be localized extracellularly with a certainty of 0.5374, as
 expected by a protein similar to the C1q complement component. In other embodiments,
 NOV11b is also likely to be localized to the microbody (peroxisome) with a certainty of
 0.1199, to the endoplasmic reticulum (membrane) with a certainty of 0.1000, and to the
 endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for
 NOV11b is between positions 15 and 16: VHS-SR.

Table 11D. Encoded NOV11b protein sequence (SEQ ID NO:30).

MVLLLLVAIPLLVHSSRGPAHYEMLGRCRMVCDPHGPRGPGPDGAPASVPPFPFGAKGEVGRGKAGLRGPP
 GPPGPRGPPGEPGRPGPPGPGPGGVAAGYVPRIAFYAGLRRPHEGYEVLRFDDVVTNVGNAYEAASG
 KFTCPMPGVYFFAYHVLMRGGDGTSMWADLMKNGQVRASAIQDADQNYDYASNSVILHLDVGDEVFIKLDG
 GKVHGGNTNKYSTFSGFIIYPD

A search of sequence databases reveals that the NOV11b amino acid sequence has 184
 of 258 amino acid residues (71%) identical to, and 198 of 258 amino acid residues (76%)
 similar to, the 258 amino acid residue C1q-related factor precursor protein from *Homo sapiens*
 (075973) ($E = 7.1 e^{-91}$). Public amino acid databases include the GenBank databases,
 SwissProt, PDB and PIR.

NOV11b is expressed in at least some of the following tissues: adrenal gland, bone
 marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra,
 brain - thalamus, brain - whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney,

lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus, right cerebellum. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources,

5 Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV11a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 11E.

Table. 11E. BLAST results for NOV11a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
Ptnr: SWISSPROT-ACC: 075973	Clq-related factor precursor - Homo sapiens	258	184/258 (71%)	198/258 (76%)	9.1e-91
ptnr:SWISSPROT- ACC:088992	Clq-related factor precursor - Mus musculus	258	156/216 (72%)	166/216 (76%)	1.8e-78
ptnr:SWISSPROT- ACC:Q9ESN4	Gliacolin precursor - Mus musculus	155	153/209 (73%)	165/209 (78%)	1.3e-77
ptnr:SWISSPROT- ACC:P02746	Complement Clq subcomponent	251	90/239 (37%)	124/239 (51%)	1.3e-29
ptnr:TREMBLNEW- ACC:AAH08983	COMPLEMENT COMPONENT 1	253	90/239 (37%)	124/239 (51%)	1.3e-29

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 11F. In the ClustalW alignment of the NOV11 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 11F. ClustalW Analysis of NOV11

- 1) NOV11a (SEQ ID NO:28)
- 2) NOV11b (SEQ ID NO:30)
- 3) ptnr: Clq-related factor precursor - Homo sapiens (SEQ ID NO:77)
- 4) ptnr: Clq-related factor precursor - Mus musculus (SEQ ID NO:78)
- 5) ptnr: Gliacolin precursor - Mus musculus (SEQ ID NO:79)
- 6) ptnr: Complement Clq subcomponent (SEQ ID NO:80)

```

NOV11a  IVLLLVANTLLVHSSRGPAHYEMLLRERMVCDIHG-PRGPPDGAIA--A-----VPPFFPPAKREVEVLCNAAL 67
NOV11b  IVLLLVANTLLVHSSRGPAHYEMLLRERMVCDIHG-PRGPPDGAIA--A-----VPPFFPPAKREVEVLCNAAL 67

```

075973	MLVIVVLIIPVLVS--GGPEGHYEMLGT--RMVCDIYP--ARGPTAGARTDGDALSEQ--GAPPPSTLVQ--EQ--KPT--TTP--	79
088992	MLVIVVLIIPVLVS--GGPDGHIYEMLGT--RMVCDIYP--ARGPTAGARSDG--DAVSEQ--GAPPPSTLVQ--EQ--KPT--TTP--	79
Q9ESN4	IVLLVLIIPVLVS--AGTSARHYEMLGT--RMVCDIYGGTKA--STAATIDR--LMQ--LPTFKQ--PK--EA--RP--PA--	73
P02746	---MQLPWGSI--FV--NL-----L--LL--LL--LDSQAQLSCTGP--AIPGI-----TP--TP--PD--QP--TP--	55
NOV11a	LRGPRP---PPGRGPPV---EPGR---GGPRGPPV--PG--VAPAGT---VPRIAFYAGLR--RPHEGYEVLRFDDVVVNV--	136
NOV11b	LRGPRP---PPGRGPPV---EPGR---GGPRGPPV--PG--VAPAGT---VPRIAFYAGLR--RPHEGYEVLRFDDVVVNV--	136
075973	PPGPPDPGPPGPGVPPPGKGEFGK---GGPPGPPG--PG--VAPAGT---VPRIAFYAGLR--RPHEGYEVLRFDDVVVNV--	156
088992	PPGPPDPGPPGPGVPPPGKGEFGK---GGPPGPPG--PG--VAPAGT---VPRIAFYAGLR--RPHEGYEVLRFDDVVVNV--	156
Q9ESN4	PPGPPDPGPPGPGVPPPGKGEFGK---GGPPGPPG--PG--VAPAGT---VPRIAFYAGLR--RPHEGYEVLRFDDVVVNV--	153
P02746	IK--EK---LPTLA--DH--EF--K--D---P--T--P--N--P--K-----V--F--G--	94
NOV11a	NAREAASGKFTCPMPGVYFFAYHVLMRGGDGTSMWADLMN--QVRASAIQAQDADQNYDYASNSVILHLHLDVGVDFEVIKLD--	216
NOV11b	NAREAASGKFTCPMPGVYFFAYHVLMRGGDGTSMWADLMN--QVRASAIQAQDADQNYDYASNSVILHLHLDVGVDFEVIKLD--	216
075973	NNDAASGKFTCPMPGVYFFAYHVLMRGGDGTSMWADLMN--QVRASAIQAQDADQNYDYASNSVILHLHLDVGVDFEVIKLD--	236
088992	NNDAASGKFTCPMPGVYFFAYHVLMRGGDGTSMWADLMN--QVRASAIQAQDADQNYDYASNSVILHLHLDVGVDFEVIKLD--	236
Q9ESN4	NNDAASGKFTCPMPGVYFFAYHVLMRGGDGTSMWADLMN--QVRASAIQAQDADQNYDYASNSVILHLHLDVGVDFEVIKLD--	233
P02746	---P-----P-----AP-----P--GPKGE--G--V--K--T--K--	120
NOV11a	---H--G--G--N--N--K--Y--S--T--F--S--G--F--I--Y--P--	238
NOV11b	---H--G--G--N--N--K--Y--S--T--F--S--G--F--I--Y--P--	238
075973	---H--G--G--N--N--K--Y--S--T--F--S--G--F--I--Y--S--	258
088992	---H--G--G--N--N--K--Y--S--T--F--S--G--F--I--Y--S--	258
Q9ESN4	---H--G--G--N--N--K--Y--S--T--F--S--G--F--I--Y--A--	255
P02746	-----	120

Tables 11E-11F list the domain descriptions from DOMAIN analysis results against NOV11. This indicates that the NOV11 sequence has properties similar to those of other proteins known to contain this domain.

Table 11E. Domain Analysis of NOV11

gnl|Smart|smart00110, C1q, Complement component C1q domain.; Globular domain found in many collagens and eponymously in complement C1q. When part of full length proteins these domains form a 'bouquet' due to the multimerization of heterotrimers. The C1q fold is similar to that of tumour necrosis factor. (SEQ ID NO:104)
CD-Length = 132 residues, 99.2% aligned
Score = 113 bits (283), Expect = 1e-26

Query: 108 PRIAFYAGL--RRPHEGYEVLRFDDVVTVNGNAYEAASGKFTCPMPGVYFFAYHVLMRGG 165
|| || || + + || + | | + + || || + || + || + +
Sbjct: 2 PRSAFVSIRSTNRPPPGQPVRFDKVLYNQGHYDPSTGKFTCPVPGVYFFSYHIESK-- 59
Query: 166 DGTSMWADLMKNGQVRASAIQAQDADQNYDYASNSVILHLHLDVGVDFEVIKLDGGKVHG-GNT 224
| ++ || || | + | | | + | | || + || + || + || +
Sbjct: 60 -GRNVKVSIMKNGIQVMRECDEYQKGLYQVASGGALLQLRQGDQVWLELDDKKNGLYAGE 118
Query: 225 NKYSTFSGFIYPD 238
|||||++||
Sbjct: 119 EVDSTFSGFLFPD 132

Table 11F. Domain Analysis of NOV11

gnl|Pfam|pfam00386, C1q, C1q domain. C1q is a subunit of the C1 enzyme complex that activates the serum complement system. (SEQ ID NO:112)
CD-Length = 125 residues, 100.0% aligned
Score = 102 bits (253), Expect = 3e-23

Query: 111 AFYAGLR-RPHEGYEVLRFDDVVTVNGNAYEAASGKFTCPMPGVYFFAYHVLMRGGDGT 169
|| | || + + || + | | + + || || + || + || + || +
Sbjct: 1 AFTAIRSTRPAPGQPVIFDEVLYNQGHYDPATGKFTCPVPGLYYFNHVSSEK---GTN 57

Query: 170 MWADLMKNGQVRASAIQDADQNYDYASNSVILHLDVGDEVFIKLDGGKVHG--GNTNKY 227
 + | | + | | | + | | | | + | | | | + + | | +
 Sbjet: 58 VCVSLMRNGVPVMSFCDEYAKGTYQVASGGAVLQLRQGDRVWLELDDKQTNGLLGGEGVH 117
 Query: 228 STFSGFTII 235
 | | | | ++
 Sbjet: 118 SVFSGFLL 125

The first component of complement system is a calcium-dependent complex of the 3 subcomponents C1q, C1r, and C1s. Subcomponent C1q binds to immunoglobulin complexes with resulting serial activation of C1r (enzyme), C1s (proenzyme) and the other 8 components of complement. It contains collagen like domains. It has been shown that fibronectin binds to C1q in the same manner that it binds collagen. A major function of the fibronectins is in the adhesion of cells to extracellular materials such as solid substrata and matrices. Because fibronectin stimulates endocytosis and promotes the clearance of particulate material from the circulation, the results suggest that fibronectin functions in the clearance of C1q-coated material such as immune complexes or cellular debris. Many examples of deficiencies of C1q have been reported, most of them associated with systemic lupus erythematosus or glomerulonephritis.

The complement system plays a paradoxical role in the development and expression of autoimmunity in humans. The activation of complement in SLE contributes to tissue injury. In contrast, inherited deficiency of classic pathway components, particularly C1q, is probably associated with the development of SLE. This leads to the hypothesis that a physiologic action of the early part of the classic pathway protects against the development of SLE and implies that C1q may play a key role in this respect. C1q-deficient (C1qa^{-/-}) mice have been shown to have increased mortality and higher titers of autoantibodies, compared with strain-matched controls. Of the C1qa^{-/-} mice, 25% have been shown to have glomerulonephritis with immune deposits and multiple apoptotic cell bodies. Among mice without glomerulonephritis, there were significantly greater numbers of glomerular apoptotic bodies in C1q-deficient mice compared with controls. The phenotype associated with C1q deficiency was modified by background genes. These findings are compatible with the hypothesis that C1q deficiency causes autoimmunity by impairment of the clearance of apoptotic cells.

The C1q-related factor is a recently discovered protein which has homology to C1q. Since this is a relatively new discovery, very little is known about its function. But conclusions could clearly be derived from its expression pattern and its homology to C1q. Based on its expression pattern it has been suggested that this protein may be involved in motor function.

The functions of C1q has been described above and include role in binding to immunoglobulin complexes, cell adhesion, autoimmunity and apoptosis, among others.

The disclosed NOV11 nucleic acid of the invention encoding a C1q-related factor-like protein includes the nucleic acid whose sequence is provided in Table 11A, 11C, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 11A or 11C while still encoding a protein that maintains its C1q-related factor-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 29 percent of the bases may be so changed.

The disclosed NOV11 protein of the invention includes the C1q-related factor-like protein whose sequence is provided in Table 11B or 11D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 11B or 11D while still encoding a protein that maintains its C1q-related factor-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 29 percent of the residues may be so changed.

The protein similarity information, expression pattern, and map location for the C1q-related factor-like protein and nucleic acid disclosed herein suggest that this C1q-related factor may have important structural and/or physiological functions characteristic of the C1q family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon.

The nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. Based on the TaqMan data, the compositions of the present invention, will have efficacy for treatment of patients suffering from: cancer of the colon, kidney, ovary, skin and brain. Since it is over expressed in cell lines derived from these tissues it can also be used as a diagnostic marker for cancer in these tissues. The expression of the novel gene of this invention upon activation of HUVEC and the homology of the novel protein of this invention to C1q may indicate that it is secreted by endothelial cells in areas of inflammation where Th1 cells are infiltrating the inflammation site such as Rheumatoid Arthritis and Inflammatory Bowel Disease. Based on its homology to C1q, the novel protein could be either pro-inflammatory activating the complement cascade and be a useful target for a monoclonal antibody to block this effect. Alternatively, this protein may act as a competitor of C1q and so act to down regulate complement mediated damage of endothelial cells. In this case it could be used as a protein therapeutic. IFN gamma also induces production of this protein by airway epithelial cell lines NCI-H292 and dermal fibroblasts indicating again that it may play a role in Th1 inflammatory diseases such as rheumatoid arthritis, multiple sclerosis, inflammatory bowel diseases and psoriasis and other diseases, disorders and conditions of the like. Because of its high homology to C1q-related factor, this novel protein may also play a role in disorders of the nervous system involved in motor function.

Based on its homology to C1q, the novel protein of invention may also play a role in the pathogenesis of systemic lupus erythematosus and glomerulonephritis and therefore could be used for detection and treatment of these diseases. Thus this protein may be involved in autoimmunity. Since the novel protein of invention has a Collagen triple helix repeat domain, it is likely that this protein may be involved in collagen related disorders and processes such as but not limited to osteogenesis, rheumatoid arthritis and osteoarthritis.

Finally, presence of somatotropin-like domain in the novel protein of invention suggests that it may have somatotropin (growth hormone) like function and behave as a growth hormone and be useful in control of growth and development/differentiation related functions such as but not limited maturation, lactation and puberty. Because of the involvement of growth hormone in many different physiologic functions, the novel protein may be involved in causing osteoporosis, obesity, aging and reproductive malfunction and hence could be used in treatment and/or diagnosis of these disorders.

The NOV11 nucleic acid, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV11 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV11 protein have multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, contemplated NOV11 epitope is from about amino acids 20 to 120. In another embodiment, the contemplated NOV11 epitope is from about amino acids 130 to 150. In further embodiments, the contemplated NOV11 epitope is from about amino acids 170 to 210, or from about amino acids 220 to 240. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV12

A disclosed NOV12 nucleic acid of 5895 nucleotides (also referred to as SC132340676_A) encoding an plexin-1-like protein is shown in Table 12A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 77-79 and ending with a TGA codon at nucleotides 5798-5800. The putative untranslated regions are underlined and are upstream from the initiation codon and downstream from the termination codon in Table 12A. The start and stop codons are in bold letters.

Table 12A. NOV12 nucleotide sequence (SEQ ID NO:31).

```

CAGGGCTGAAGCTCCTGGCACCATGATGCTCACCCAGCAGGACCAGAGCACCAGGGCCCAAGGCCCCAGCCTGCCATGC
CGCTGCCACCCGCGGAGCCTGCAGGTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCCGGGCATGTGGGCTGAGGCA
GGCTTGCCACGGGCAGGCGGGGGTTACAGCCCCCTTCCGCACCTTCTCGGCCAGCGACTGGGGCCTCACCCACCTAGT
GGTGCCATGAGCAGACAGGCGAGGTGTATGTGGGCGCAGTGAACCGCATCTATAAGCTGTGGGGAACTGACACTGCTGC
GGGCCCCAGCTCACGGGCCCTGTGGAGGACAACGAGAAGTGCTACCCGCGCCCGAGCGTGCACTCCTGCCCCACGGCCTG
GGCAGTACTGACAACGTCAACAAGCTGCTGCTGCTGGACTATGCCGTAACCGCCTGCTGGCCTGTGGCAGCGCCTCCCA
GGGCATCTGCCAGTTCTGCGTCTGGACGATCTCTTCAAAGTGGGTGAGCCACACCACCGTAAGGAGCACTACCTGTCCA
CGCTGCAGGAGCGCAGGCGAGCACTTCTTACGTCCTCAAGATCGTGCGGCTCTGTGTGGAGCAGCCCAATTTCTACTCGTA
ACACCCATCGATGGCAAGTCCGAGTACTTCCCCACACTGTCCAGCCGTCGGCTCATGGCCAACGAGGAGGATGCCGACAT
GTTCCGCTTCTGTACAGGATGAGTTTGTGTGTCATCACAGCTCAAGATCCCTTCGGACAGCGTGTCCAAGTTCCCGGCCT
TTGACATCTACTATGTGTACAGCTTCCGACGAGCAGTTTGTCTACTACCTCAGCTGCAGCTAGACACACAGCTGACC
TCGCCCTGATGCCGCCGCGAGCACTTCTTACGTCCTCAAGATCGTGCGGCTCTGTGTGGAGCAGCCCAATTTCTACTCGTA
CGTTGAGTTCCCCATTGGCTGCGAGCAGGCGGGTGTGGAGTACCGCCTGGTGCAGGATGCCCTACCTGAGCGCGCCGGCC
GTGCCCTGGCCACCAGCTGGGCCTGGCTGAGGACGAGGACGTGCTGTCACTGTGTTGCGCCAGGGCCAGAAGAACCGC
GTGAAGCCACCAAGGAGTCAGCACTGTGCCTGTTACGCTCAGGGCCATCAAGGAGAAGATTAAGGAGCGCATCCAGTC
CTGCTACCGTGGTGAGGGCAAGCTCTCCCTGCCGTGGTGCTCAACAAGGAGCTGGGCTGCATCACTCGCCCTGCAGA

```


TCGATGACGACTTCTGCGGGCAGGACTTCAACCAGCCCCTGGGGGGCACAGTCACCATTGAGGGGACGCCCCGTGTCGTG
 GACAAGGATGATGGCCTGACCGCCGTGGCTGCCTATGACTATCGGGGCCGCACTGTGGTATTGCGCCGGCACGCGAAGTGG
 CCGCATCCGCAAGATCCTGGTGGACCTCTCAAACCCGGTGGCCGGCTGCCCTGGCTACGAGAGCGTGTGGCCGAGG
 AGGCGAGCCCCATCCTGCGAGACCTCGTCTCAGCCCCAACACCAAGTACCTCTACGCCATGACCGAGAAGCAGGTGACG
 CGGGTGCCTGTGGAGAGCTGTGTGCAGTACACGTCCTGTGAGCTGTGTCTGGGGTCACGGGACCCCCACTGTGGCTGGTG
 TGTCTGCACAGCATGTGCTGCGCGCGGAGCGCCTGTGAGCGAGCAGACGAGCCCCAGCGCTTGTCTGCGACCTGCTGC
 AGTGTGTGCAGCTGACTGTGCAGCCCCGCAATGTGTCTGTCAACATGTCCAGGTCCCAGTACTTGTGCTGCAGGCCTGG
 AACGTGCCGTGACCTCTCAGTGGCGTCAACTGCTCCTTCGAGGACTTACGGAATCTGAGAGCGTCCCTGGAGGATGGCCG
 GATCCACTGCGCTAAAACCCATCACCTGGCGCGGAGGTGGCGCCATCACGCGGGCCAGGGTGAGGGAGACCAGCGGTGGTGA
 AACTCTACCTAAAGTCCAAGGAGACAGGGAAGAAGTTGCGTCTGTGGACTTCGTCTTCTACAACTGCAGCGTCCACCAG
 TCGAGCTGCTGTCTGTGTCAACGGCTCCTTTCCCTGCCACTGGTGCAAATACCGCCACGTGTGCACACACACAGTGGC
 TGACTGCGCCTTCTGGAGGGCGGTGTCAACGTGTCTGAGGACTGCCACAGATCCTGCCCTCCACGCAGATCTACGTGC
 CAGTGGGAGTGGTAAAACCCATCACCTGGCGCGGAGGTGGCGCCATCACGCGGGCCAGGGTGAGGGAGACCAGCGGTGGTGA
 CTCTTCCACATCCCGGGCAGCCCGGGCCGTGTCAACGGCTGCGCTTCAACAGCTCCAGCCTGCAGTGCCAGAATTCTCTC
 GTACTCCTACGAGGGGAACGATGTGAGCGACCTGCCAGTGAACCTGTGAGTGTGTGGAACGGCAACTTGTGCTATTGACA
 ACCACAGAACATCCAGGCGCACCTCTACAAGTGGCCGGCCCTGCGCGAGAGCTGCGGCCCTGTGCTCAAGGCCGACCCG
 CGCTTCGAGTGGATGGTGTGCGTGGCGGAGCGCGCTGCTCACCACACTACCGCGCCCTGTCAACCAAGCGCTTCACT
 GATGCACGCGCGTCAAGGCGAGCTGCGTGCACCGACCCCAAGATCCTCAAGCTGTCCCCGAGACGGGCCGAGGCAGG
 GCGGCACGCGGCTCACTATCACAGCGAGAACCTGGGCTGCGATTGGAAGAGTGTGCTGTGGCGTGTGCGTGGGCAAG
 GTGCTGTGCAGCCCTGTGGAGAGCGAGTACATCAGTGGCGAGCAGATCGTCTGTGAGATCGGGGACGCCAGCTCCGTGCG
 TGCCCATCGAGCTTGTGGTGTGTGTGCGGGAGCTGCTCACCACACTACCGCGCCCTGTCAACCAAGCGCTTCACT
 TCGTGACACCAACTTCTACCGTGTGAGCCCTCCCGTGGGCTGTGTGAGGGGCACTGGATGGCATCGAGGGAAGC
 CACTGAACGCGAGGAGTGTGTGGTGTGTGCGTGGTGGCGGCCCTGTCTCTCTGCTCCAGGAGGAACCTCCCG
 TGAGATCCGGTGCCTGACACCCCCGGGCGAGCCCTGGCAGCGCTCCCATCATCAACATCAACCGCGCCGAGCTCA
 CCAACCTGTAGGTGAAGTACAACCTACACCGAGGACCCCACTCCTGAGGATCGACCCGAGTGGAGCATCAACAGCGGT
 GGACCCCTCCTGACCGTCAAGGCAACCTGGCCACTGTGCGTGAACCCGAATCCGGGCCAAGATGAGGAGGCACTGA
 GAGGAGAACTGCCTGGTGTACAATGACACCACATGGTATGCCGCGCCCGTGTGTGGCCAAACCTGTGCGCAGCCAC
 CAGAGCTGGGGGAGCGGCCGATGAGCTGGGCTTCGTGATGACAACTGCGCTCCCTGCTTGTGCTCAACTCCACCTCC
 TTCCTCTACTACCTGACCCCGTACTGGAGCCACTCAGCCCACTGGCCTGTGAGCTGAAGCCAGCTCCCCACTCAT
 CTTCAAGGCGCGGAACCTTGTCCACTGACCCCGCAACTCCGACTCAACTACAGGTGCTCATCGGCTCCACACCT
 GTACCTCACCCTGTGCGAGACGCAACTGCTGTGCGAGGCGCCCACTCACTGGGCAGCACAAGGTACGGTGTGCTGCA
 GGTGGCTTCGAGTTCTCGCCAGGGACACTGCAGGTGACTCGGACAGCCTGTGACGCTGCCATGTGTGGCATTGG
 CCGAGGCGGGGCTCTCTGTGTGTGTCATCGTGGCTGTGCTCATCGCTACAAGCGCAAGTACAGAGATGCTGACCGCA
 CTTCAAGGCGCGTGCAGCTCAGATGGACAACCTGGAGTCCGCGTGGCCCTCGAATGCAAGGAAGCTTTGACAGAGCTG
 CAGACAGACATCCAGAGCTGACCAATGACCTGGACGGTGGCGCATCCCCCTTCTTACTACCGGACATATGCCATGCG
 GGTGCTCTTCTGGGATCGAGGACACCTGTGCTCAAGGAGATGGAGGTACAGGCCAATGTGGAGAAGTGTGCTGACAC
 TGTTCGGGAGCTGCTGACCAAGAAGCACTTCTGTGACCTTCATCGCAGCTGGAGGACAGCGAGCTTCTCCATG
 CGGACCGCGGGAATGTGGCTCGCTCATGACGCCCCGAGGCGGAGATGGAATACGCCACAGGCGGTGTCAAGCA
 GCTGCTTTCCGACCTCATCGAGAAGAACCTGGAGAGCAAGAACCACCCCAAGCTGCTACTGCGCCGCGCAACTGAGTCCG
 TGGCAGAGAAGATGCTAACTAAGTTCACCTTCTTGTATAAGTTCTCAAGGAGTGGCTGGGGAGCCGCTGTTC
 ATGCTGTACTGCGCCATCAGCAGCAGATGGAGAAGGGCCCCATTGACGCCATCAGCGGTGAGGCACGCTACTCCCTGAG
 TGAGGACAAGCTCATCCGCGCAGCAGATTGACTACAAGACACTGACCTGAAGTGTGTAACCTGAGAATGAGAATGCAC
 CTGAGGTGCGGCTGAGGGGCTGGACTGTGACACGGTCAACCGGCAAGGAGAAGCTGCTGGACGCTGCGCTCAAGGGC
 GTGCCCTACTCCAGCGGCCCAAGGCCGCGACATGGACCTGGAGTGGCGCCAGGGCCGATGGCGCGCATCATCCTGCA
 GGACGAGGACGTCAACACCAAGATTGACAACGATTGGAAGAGGCTGAACACACTGGCTCACTACCAGGTGACAGACGGGT
 CCTCGGTGGCATGCTGGTCCCAAGCAGACGTCGCGCTACAACATCTCCAACCTCCTCCACCTTCAACCAAGTCCCTCAGCAGA
 TACGAGAGCATGCTGCGCAGCGCCAGCAGCCCCGACAGCCTGCGCTCGCGCAGCGCCATGATCAGCCCCGACCTGGAGAG
 CGGCACCAAGCTGTGGCAGCTGGTGAAGAACCACGACCACTGGACCAAGCTGAGGGTGACCGCGGCAAGATGGTCT
 CGGAGATCTACTTGACACGGCTACTGGCCACCAAGCAGGGCACACTGCAGAAGTTTGTGGACGACCTTTTGAGACCATC
 TTCAGCACGGCACACCGGGGCTCAGCCCTGCCGCTGGCCATCAAGTACATGTTGCACTTCTGGATGAGCAGGCCGACAA
 GCACAGATCCACAGTGTGACGTGCGCCACACTGGAAGAGCAACTGCAGCCTGCCCTGCGCTTCTGGGTGAACGTGA
 TCAAGAACCACAGTTTGTGTTGACATTCACAAGAACAGCATCAGGACGCGCTGCTGTGCGTGGTGGCCGACCTTC
 ATGGAATCCTGCTCCACCTGTGACACAAGCTGGGCAAGGACTCACCTCCAACAAGCTGCTTACGCCAAGGACATCCC
 CAATACAGAGCTGGGTGGAGAGGAGTACTATGCAGACATCGCCAAGATGCCAGCATCAGCGACCAAGGACATGAGTG
 CGTATCTGGCTGAGCAGTCCCGCTGCACCTGAGCCAGTTCAACAGCATGAGCGCCTTGCACGAGATCTACTCTACATC
 ACCAAGTACAAGGATGAGGTGAGATCCTGGCAGCCCTGGAGAAGGATGAGCAGGCGCGGCGGACGCGCTGCGGAGCAA
 GCTGGAGCAGGTGGTGGACAGATGGCCCTGAGCAGCTGAGCCCCAGCTGTGATCATCCAGCATGATGCAGCGTGAGGAC
 AGCTGAGCAGGACCGGGCAGCCCTCACGCATGCGTGTGGAGTGTCCGGTGGT

In a search of public sequence databases, the NOV12 nucleic acid sequence, located on chromosome 8 has 2950 of 3362 bases (87%) identical to a plexin-1 mRNA from *Mus musculus*, (GENBANK-ID: D86948) (E = 0.0). Public nucleotide databases include all

5 GenBank databases and the GeneSeq patent database.

The disclosed NOV12 polypeptide (SEQ ID NO:32) encoded by SEQ ID NO:31 has 1925 amino acid residues and is presented in Table 12B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV12 contains a signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.6000. In other

5 embodiments, NOV12 is likely to be localized to the Golgi body with a certainty of 0.4000, to the endoplasmic reticulum (membrane) with a certainty of 0.1000, or to the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV12 is between positions 44 and 45: MWA-EA.

Table 12B. Encoded NOV12 protein sequence (SEQ ID NO:32).

```

MMLTPAGPEHRGPRPQAMPPLPPRSLOVLLLLLLLLLLLLLPGMWAEAGLPAGGGSQPPFRTFSASDWGLTHL
VVEHQTGEVYVGAVNRIYKLSGNLTLLRAHVTGPVEDNEKCYPPPSVQSCPHGLGSTDNVNKLLLLDYAANR
LLACGSASQGICQFLRLDDLFKLGEPHHRKEHYLSSVQAGSMAGVLIAGPPGQQAFLVFGTPIDGKSEYF
PTLSSRRRLMANEEDADMFGFVYQDEFVSSQLKIPSDTLSEKFAFDIYYVYSFRSEQFVYYLTQLDQLTQTSF
DAAGEHFTSKIVRLCVDDPKFYSYVEFFIGCEQAGVEYRLVQDAYLSRPGRALAHQLGLAEDEDVLETFFA
QQQKNRVKPPKESALCLFTLRAIKEKIKERIQSCYRGEGLSLPWLNLKELGCINSPLQIDDDFCGQDFNQF
LGGTVTIEGTPLFVDKDDGLTAVAAVDYRGRTVVFAGTRSGRIRKILVDLSNPGGRPALAYESVVAQEGSPI
LRDLVLSFNHQYLYAMTEKQVTRVPVESCVOYTSCELCGSRDPHCGWCVLHSMCRRDACERADPEPQFAA
DLLQCQVLTQVQPRNVSVTMSQVPVLVLQAWNVPDLSAGVNCSEDFTESESVLEDGRIHCRSPSAREVAPIT
RGQEGEQDQVVKLYLKSKEGKFAVDFVFNCSVHQSSCLSCVNGSFPCWCKYRHVCTHNVADCAFLEG
RVNVSEDCPQILPSTQIYVPGVVKPITLAARNLPQPSQSGQRYECLFHIPGSPARVTALRENSSSLQCCNS
SYSYEGNDVSDLPVNLVSVWNGNFVIDNPQNIQAHLYKCPALRESCGLCKADPRFECGWCVAERRCSLRHH
CAADTPASWMMHARHGSSRCTDPKILKLSPETGPRQGGTRLTITGENLGLREFDVLGVRVGKVLCSFVESEY
ISAEQIVCEIGDASSVRAHDALVEVCVRDCSPHYRALS PKRFTFVTPTFYRVSPSRGPLSGGTWIGIEGSHL
NAGSDVAVSVGGRPCSFWSRRNSREIRCLTPPGQSPGSAPIIININRAQLTNPEVKYNYTEDPTILRIDPE
WSINSGGTLTTLVTGTNLATVREPRIRAKYGGIERENCLVYNDTMMVCRAPSVANPVRSPPELGERPDELGFV
MDNVRSLVLNLSSTFELYPDVLEPLSPTGLLELKPSSPLILKGRNLLPPAPGNSRLNYTVLIGSTPCTLT
SETQLLCEAPNLTGQHKVTVRAGGEFFSPGTLQVYSDSLTLPAIVGIGGGGGLLLIVIVAVLIAVKRKS
ADRTLKRLQLQMDNLESVALECKEAFELQTDIHELTNDLDGAGIPFLDYRTYAMRVLPFGIEDHPVLKEM
EVQANVEKSLTLFGQLLTKKHFLFTFIRTLAQRSFSMRDRGNVASLIMTALQGEMEYATGVKQLLSDLIE
KNLESKNHPKLLLRPTESVAEKMLTNWTFLLYKFLKECAGEPLFMYCAIKQQMEKGPIDAITGEARYSL
SEDKLIRQQIDYKTLTLNCVNPENENAPEVVPKGLDCDVTVOAKEKLLDAAYKGVYPSQRPKAADMDEWRQ
GRMARIILQDEDVTTKIDNDWKRNLTLAHYQVTDGSSVALVPKQTSAYNISNSSTFTKLSRYESMLRTASS
PDSLRSTPMTIPDLESQTKLWHLVKNHDLQREGDRGSKMVSEIYLTRLLATKQGTLOKEFVDDLEETIFS
TAHRGSALPLAIKYMFDLDEQADKHQIHDADVHTWKSNCSLPLRFVNVIKNPQFVFDIHKNSITDACLS
VVAQTMDSCSTSEHKLKGDSPENKLLYAKDIPNYKSWVERRYADIAKMPAISDQDMSAYLAEQSRHLHSQ
FNSMSALHEIYSYITKYKDEVQIILAALKDEQARRQLRSKLEQVVDTMALSS

```

A search of sequence databases reveals that the NOV12 amino acid sequence has 1820 of 1907 amino acid residues (95%) identical to, and 1859 of 1907 amino acid residues (97%) similar to, the 1894 amino acid residue plexin-1 protein from *Mus musculus* (P70206) (E = 0.0). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV12 is expressed in at least the following tissues: whole organism, brain, testis, trabecular Bone, lymph, germinal center B cells. In addition, NOV12 is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: acc:AI255192) a closely related plexin-1 homolog in species *Mus musculus*: brain, testis.

The disclosed NOV12 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 12C.

Table 12C. BLAST results for NOV12

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
Ptnr:SPTREMBL- ACC:P70206	PLEXIN 1 - Mus musculus	1894	1820/1907 (95%)	1859/1907 (97%)	0.0
ptnr:SPTREMBL- ACC:Q9UIW2	NOV/PLEXIN-A1 PROTEIN - Homo sapiens	1754	1743/1762 (98%)	1746/1762 (99%)	0.0
ptnr:SPTREMBL- ACC:Q91823	PLEXIN PRECURSOR - Xenopus laevis	1905	1603/1893 (84%)	1730/1893 (91%)	0.0
ptnr:SWISSPROT- ACC:P51805	Plexin A3 precursor (Plexin 4)	1871	1252/1874 (66%)	1483/1874 (79%)	0.0
ptnr:SPTREMBL- ACC:P70208	PLEXIN 3 - Mus musculus	1872	1245/1874 (66%)	1478/1874 (78%)	0.0

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 12D. In the ClustalW alignment of the NOV12 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 12D. ClustalW Analysis of NOV12

- 1) NOV12 (SEQ ID NO:32)
- 2) ptnr: PLEXIN 1 - Mus musculus (SEQ ID NO:81)
- 3) ptnr: NOV/PLEXIN-A1 PROTEIN - Homo sapiens (SEQ ID NO:82)
- 4) ptnr: PLEXIN PRECURSOR - Xenopus laevis (SEQ ID NO:83)
- 5) ptnr: Plexin A3 precursor (Plexin 4) (SEQ ID NO:84)
- 6)

NOV12	MMLTPAGPEHRGPRPQAMP	PPRSLOVLLLL	LLLL	LLLP	MAEA	LRAGG	SSQPPFRTFS	SANLWGLTHLVVHEQTGE	80
P70206	-----	MPLPLSSRTLL	LLLL	LL	RAVNT	YSS	FAGL	EDPAFRTFVARDWGLTHLVVHEQTGE	60
Q9UIW2	-----	-----	-----	-----	MAEA	LRAGG	SSQPPFRTFS	SGSLWGLTHLVVHEQTGE	40
Q91823	-----	MLLHAERPLPFF	WTET	VLL	SS	FAAT	-----	EDYPKDFRTHTGSLWGLTHLVVHEQTGE	55
P51805	-----	-----	-----	-----	-----	-----	-----	-----	44
NOV12	VYVGAVNRIYKLS	NTLLRAHVTGPVEI	NEKCYPPPSVQSCPHGL	STIN	NNKLL	LLDYA	NNKLLACGSASQGICQFLR	160	
P70206	VYVGAVNRIYKLS	NTLLRAHVTGPVEI	NEKCYPPPSVQSCPHGL	STIN	NNKLL	LLDYA	NNKLLACGSASQGICQFLR	140	
Q9UIW2	VYVGAVNRIYKLS	NTLLRAHVTGPVEI	NEKCYPPPSVQSCPHGL	STIN	NNKLL	LLDYA	NNKLLACGSASQGICQFLR	120	
Q91823	VYVGAVNRIYKLS	NTLLRAHVTGPVEI	NEKCYPPPSVQSCPHGL	STIN	NNKLL	LLDYA	NNKLLACGSASQGICQFLR	135	
P51805	VYVGAVNRIYKLS	NTLLRAHVTGPVEI	NEKCYPPPSVQSCPHGL	STIN	NNKLL	LLDYA	NNKLLACGSASQGICQFLR	124	
NOV12	LDLFLKLGEPHHRKEHY	LSSVQEAGSMAGVLI	AGPPGQQA	KLFGVT	PIDGKSEYFPTLSSRRLM	NEEDALHFGVYQI	240		
P70206	LDLFLKLGEPHHRKEHY	LSSVQEAGSMAGVLI	AGPPGQQA	KLFGVT	PIDGKSEYFPTLSSRRLM	NEEDALHFGVYQI	220		
Q9UIW2	LDLFLKLGEPHHRKEHY	LSSVQEAGSMAGVLI	AGPPGQQA	KLFGVT	PIDGKSEYFPTLSSRRLM	NEEDALHFGVYQI	200		
Q91823	LDLFLKLGEPHHRKEHY	LSSVQEAGSMAGVLI	AGPPGQQA	KLFGVT	PIDGKSEYFPTLSSRRLM	NEEDALHFGVYQI	160		
P51805	LDLFLKLGEPHHRKEHY	LSSVQEAGSMAGVLI	AGPPGQQA	KLFGVT	PIDGKSEYFPTLSSRRLM	NEEDALHFGVYQI	202		
NOV12	EFVSSQLKIPSDT	LSKFPADFII	YVYSERSEQ	FVYLLT	QLD	TQLTSPDAAGEHFFTSKIVRLC	DIIPKFSYVEFPICQ	320	
P70206	EFVSSQLKIPSDT	LSKFPADFII	YVYSERSEQ	FVYLLT	QLD	TQLTSPDAAGEHFFTSKIVRLC	DIIPKFSYVEFPICQ	300	
Q9UIW2	EFVSSQLKIPSDT	LSKFPADFII	YVYSERSEQ	FVYLLT	QLD	TQLTSPDAAGEHFFTSKIVRLC	DIIPKFSYVEFPICQ	280	
Q91823	EFVSSQLKIPSDT	LSKFPADFII	YVYSERSEQ	FVYLLT	QLD	TQLTSPDAAGEHFFTSKIVRLC	DIIPKFSYVEFPICQ	160	
P51805	EFVSSQLKIPSDT	LSKFPADFII	YVYSERSEQ	FVYLLT	QLD	TQLTSPDAAGEHFFTSKIVRLC	DIIPKFSYVEFPICQ	282	

5	NOV12	EQAGVEYRLVODAYLSRPGALAHJGLAEDEVLFTVFAQGOINRVKPPKESALCLFTLRATKEKIKERIOSCYRGGH	400
	P70206	EQAGVEYRLVODAYLSRPGALAHJGLAEDEVLFTVFAQGOINRVKPPKESALCLFTLRATKEKIKERIOSCYRGGH	380
	Q9UIW2	EQAGVEYRLVODAYLSRPGALAHJGLAEDEVLFTVFAQGOINRVKPPKESALCLFTLRATKEKIKERIOSCYRGGH	360
	Q91823	EQAGVEYRLVODAYLSRPGALAHJGLAEDEVLFTVFAQGOINRVKPPKESALCLFTLRATKEKIKERIOSCYRGGH	160
10	P51805	SWRVEYRIVASHLAKTLLLAQALVPAEDVLFTHSGQINASIRTOTILCLFTISNAHRRRTUQYRGGT	362
	NOV12	LSLPWLLNKLGCINSPLQIDDDFCQDHNPLGGTVTIEGTPLFVDREDGLTAVAAVDYRCRTVVFAGTSGRIKRLIV	480
	P70206	LSLPWLLNKLGCINSPLQIDDDFCQDHNPLGGTVTIEGTPLFVDREDGLTAVAAVDYRCRTVVFAGTSGRIKRLIV	460
	Q9UIW2	LSLPWLLNKLGCINSPLQIDDDFCQDHNPLGGTVTIEGTPLFVDREDGLTAVAAVDYRCRTVVFAGTSGRIKRLIV	440
15	Q91823	LSLPWLLNKLGCINSPLQIDDDFCQDHNPLGGTVTIEGTPLFVDREDGLTAVAAVDYRCRTVVFAGTSGRIKRLIV	160
	P51805	LAIPWLLNKLGCINSPLQIDDDFCQDHNPLGGTVTIEGTPLFVDREDGLTAVAAVDYRCRTVVFAGTSGRIKRLIV	442
20	NOV12	DIISNIGRPAAYESVVAQEGSPILKDLVLSINQIYLYAMTEKQVTRVPVESCVOYTSCELCGSKIPHCQWVLEHSC	560
	P70206	DIISNIGRPAAYESVVAQEGSPILKDLVLSINQIYLYAMTEKQVTRVPVESCVOYTSCELCGSKIPHCQWVLEHSC	540
	Q9UIW2	DIISNIGRPAAYESVVAQEGSPILKDLVLSINQIYLYAMTEKQVTRVPVESCVOYTSCELCGSKIPHCQWVLEHSC	520
	Q91823	DIISNIGRPAAYESVVAQEGSPILKDLVLSINQIYLYAMTEKQVTRVPVESCVOYTSCELCGSKIPHCQWVLEHSC	160
25	P51805	IG---GQDHLTFPPVVDSPILKDLVLSINQIYLYAMTEKQVTRVPVESCVOYTSCELCGSKIPHCQWVLEHSC	518
	NOV12	PRDACERADPDPFADLLQCVLTQVPPNVSVTMSQVILVIOQWNPDLGAGVNSFEEDFTESVLLDGRHCRST	639
	P70206	PRDACERADPDPFADLLQCVLTQVPPNVSVTMSQVILVIOQWNPDLGAGVNSFEEDFTESVLLDGRHCRST	618
	Q9UIW2	PRDACERADPDPFADLLQCVLTQVPPNVSVTMSQVILVIOQWNPDLGAGVNSFEEDFTESVLLDGRHCRST	598
30	Q91823	PRDACERADPDPFADLLQCVLTQVPPNVSVTMSQVILVIOQWNPDLGAGVNSFEEDFTESVLLDGRHCRST	160
	P51805	PGVAGLGSALHGFADPSPKIVVVRIRINNSVSPGQITVTLHNVPDLGAGVNSFEEDFTESVLLDGRHCRST	597
35	NOV12	SAREVAPITRGCGDORVVKLYLKSKEGKKEASVDFEYNSVHQSCLACVNSFPCHWCKYRHVCNNADCAFL	719
	P70206	SAREVAPITRGCGDORVVKLYLKSKEGKKEASVDFEYNSVHQSCLACVNSFPCHWCKYRHVCNNADCAFL	695
	Q9UIW2	SAREVAPITRGCGDORVVKLYLKSKEGKKEASVDFEYNSVHQSCLACVNSFPCHWCKYRHVCNNADCAFL	675
	Q91823	SAREVAPITRGCGDORVVKLYLKSKEGKKEASVDFEYNSVHQSCLACVNSFPCHWCKYRHVCNNADCAFL	160
40	P51805	SLQIRALTRH--ATITTRIQILKKEGKKEASVDFEYNSVHQSCLACVNSFPCHWCKYRHVCNNADCAFL	674
	NOV12	GRVNVREDCPQILPSTQIVPVGVVKPITLAAHLEQPOSQGRGYECLEHFGSPARVTLKINSSLOQNNSSYSYEN	799
	P70206	GRVNVREDCPQILPSTQIVPVGVVKPITLAAHLEQPOSQGRGYECLEHFGSPARVTLKINSSLOQNNSSYSYEN	775
	Q9UIW2	GRVNVREDCPQILPSTQIVPVGVVKPITLAAHLEQPOSQGRGYECLEHFGSPARVTLKINSSLOQNNSSYSYEN	755
45	Q91823	GRVNVREDCPQILPSTQIVPVGVVKPITLAAHLEQPOSQGRGYECLEHFGSPARVTLKINSSLOQNNSSYSYEN	160
	P51805	GNVHSPGCTEILPGDGLIIVGVVKPITLAAHLEQPOSQGRGYECLEHFGSPARVTLKINSSLOQNNSSYSYEN	754
50	NOV12	DVSDLEVNLSVVNNNEVILNINQIAHLYKCPALPESGGLCKADPREFECGNCVAERCCSLRHHPATIPASMMHARH	879
	P70206	DVSDLEVNLSVVNNNEVILNINQIAHLYKCPALPESGGLCKADPREFECGNCVAERCCSLRHHPATIPASMMHARH	855
	Q9UIW2	DVSDLEVNLSVVNNNEVILNINQIAHLYKCPALPESGGLCKADPREFECGNCVAERCCSLRHHPATIPASMMHARH	835
	Q91823	DVSDLEVNLSVVNNNEVILNINQIAHLYKCPALPESGGLCKADPREFECGNCVAERCCSLRHHPATIPASMMHARH	160
55	P51805	EHGTEYDEFAVVDDEPILKPSFRLYKQWQIPVGLCKADPREFECGNCVAERCCSLRHHPATIPASMMHARH	833
	NOV12	SSRCTDPKILKLSPETGPRQGGTRLTITGNNGLRFEEDVRLGVHGVKVLCSFVESEYISAEQIVCEIGDAITPAHDALV	959
	P70206	SSRCTDPKILKLSPETGPRQGGTRLTITGNNGLRFEEDVRLGVHGVKVLCSFVESEYISAEQIVCEIGDAITPAHDALV	935
	Q9UIW2	SSRCTDPKILKLSPETGPRQGGTRLTITGNNGLRFEEDVRLGVHGVKVLCSFVESEYISAEQIVCEIGDAITPAHDALV	915
60	Q91823	SSRCTDPKILKLSPETGPRQGGTRLTITGNNGLRFEEDVRLGVHGVKVLCSFVESEYISAEQIVCEIGDAITPAHDALV	160
	P51805	GIVSHSSTQTHLVNKEKTHVTIVNHNLSREGRAG--RINSTPAEVIASRIVINDESLVPSPPPGH	911
65	NOV12	FVCRDRCSPHYRALSFKRFTFTPTFYRVSFSGGPISGGTWIGIEGSHNAGSDVAVSVGGRCPSFUSRNRREIRCIY	1039
	P70206	FVCRDRCSPHYRALSFKRFTFTPTFYRVSFSGGPISGGTWIGIEGSHNAGSDVAVSVGGRCPSFUSRNRREIRCIY	1013
	Q9UIW2	FVCRDRCSPHYRALSFKRFTFTPTFYRVSFSGGPISGGTWIGIEGSHNAGSDVAVSVGGRCPSFUSRNRREIRCIY	993
	Q91823	FVCRDRCSPHYRALSFKRFTFTPTFYRVSFSGGPISGGTWIGIEGSHNAGSDVAVSVGGRCPSFUSRNRREIRCIY	160
70	P51805	ILNNGCMADEFTQEQVSEVWPTTQGSFPTIAYTRITNSNDVNRITIRDSEQIVR--DAKAVITG	989
	NOV12	PPQOSPGSAPIININRAQIINFEVKNYTEDPTILRIDPEWSINSGGTLTGTGNIATVREPRIRAKYGGIERENS	1117
	P70206	PPQOSPGSAPIININRAQIINFEVKNYTEDPTILRIDPEWSINSGGTLTGTGNIATVREPRIRAKYGGIERENS	1092
	Q9UIW2	PPQOSPGSAPIININRAQIINFEVKNYTEDPTILRIDPEWSINSGGTLTGTGNIATVREPRIRAKYGGIERENS	1072
75	Q91823	PPQOSPGSAPIININRAQIINFEVKNYTEDPTILRIDPEWSINSGGTLTGTGNIATVREPRIRAKYGGIERENS	160
	P51805	ILSTIGSQDPTILADKANISSGLITITQNTITLSTISINSLATVSTHLLVQMTVAFRHTTNT	1069
80	NOV12	LYVNTTMYCRAPVAVVRSPPELGERPDELFTVHNVSLVINTSFLYYPDPVLEPLSPTGLLEKPSPLILKRG	1197
	P70206	LYVNTTMYCRAPVAVVRSPPELGERPDELFTVHNVSLVINTSFLYYPDPVLEPLSPTGLLEKPSPLILKRG	1172
	Q9UIW2	LYVNTTMYCRAPVAVVRSPPELGERPDELFTVHNVSLVINTSFLYYPDPVLEPLSPTGLLEKPSPLILKRG	1152
	Q91823	LYVNTTMYCRAPVAVVRSPPELGERPDELFTVHNVSLVINTSFLYYPDPVLEPLSPTGLLEKPSPLILKRG	160
85	P51805	QINTITAILKAGGFLGRPCRAQTHPTFELLHQTARSNRSEITYPISFEPIGSLVDVTCGVVPTK	1149
	NOV12	NLLPPAPGNSRIINYTLIGSTPCLTVSETOLLCEARNLTGQHKVTVRAGGEFSPGLQVYSDSLTLPALVIGIGGGG	1277
	P70206	NLLPPAPGNSRIINYTLIGSTPCLTVSETOLLCEARNLTGQHKVTVRAGGEFSPGLQVYSDSLTLPALVIGIGGGG	1252
	Q9UIW2	NLLPPAPGNSRIINYTLIGSTPCLTVSETOLLCEARNLTGQHKVTVRAGGEFSPGLQVYSDSLTLPALVIGIGGGG	1232
90	Q91823	NLLPPAPGNSRIINYTLIGSTPCLTVSETOLLCEARNLTGQHKVTVRAGGEFSPGLQVYSDSLTLPALVIGIGGGG	160
	P51805	NILHAALASRIINYTLIGSTPCLTVSETOLLCEARNLTGQHKVTVRAGGEFSPGLQVYSDSLTLPALVIGIGGGG	1229
95	NOV12	LLLLVIVAVLIAYKKRSRDADRTLKRLOQMNLERSVALECKEAFABLOTDIHELTDOLDGAGIPFLDYRTYAMRVLF	1357
	P70206	LLLLVIVAVLIAYKKRSRDADRTLKRLOQMNLERSVALECKEAFABLOTDIHELTDOLDGAGIPFLDYRTYAMRVLF	1332
	Q9UIW2	LLLLVIVAVLIAYKKRSRDADRTLKRLOQMNLERSVALECKEAFABLOTDIHELTDOLDGAGIPFLDYRTYAMRVLF	1312
	Q91823	LLLLVIVAVLIAYKKRSRDADRTLKRLOQMNLERSVALECKEAFABLOTDIHELTDOLDGAGIPFLDYRTYAMRVLF	160
100	P51805	LLLLVIVAVLIAYKKRSRDADRTLKRLOQMNLERSVALECKEAFABLOTDIHELTDOLDGAGIPFLDYRTYAMRVLF	1309
	NOV12	LTEDHPVLKMEVQANVEKSLTFLGOLLTKKHLLTFTIRTEAQRFSMRDRGNVASLIMTALQEMEYATGVLRQLLS	1437
	P70206	LTEDHPVLKMEVQANVEKSLTFLGOLLTKKHLLTFTIRTEAQRFSMRDRGNVASLIMTALQEMEYATGVLRQLLS	1412
	Q9UIW2	LTEDHPVLKMEVQANVEKSLTFLGOLLTKKHLLTFTIRTEAQRFSMRDRGNVASLIMTALQEMEYATGVLRQLLS	1392
	Q91823	LTEDHPVLKMEVQANVEKSLTFLGOLLTKKHLLTFTIRTEAQRFSMRDRGNVASLIMTALQEMEYATGVLRQLLS	160
105	P51805	LTEDHPVLKMEVQANVEKSLTFLGOLLTKKHLLTFTIRTEAQRFSMRDRGNVASLIMTALQEMEYATGVLRQLLS	1389
	NOV12	LTERNLESNHPKILLRPESVAEKMLTNVFTFLYKFLKECAGPLFMLYCAIKQMEKGPIDAITGEARYSLSEDKI	1517
	P70206	LTERNLESNHPKILLRPESVAEKMLTNVFTFLYKFLKECAGPLFMLYCAIKQMEKGPIDAITGEARYSLSEDKI	1491
	Q9UIW2	LTERNLESNHPKILLRPESVAEKMLTNVFTFLYKFLKECAGPLFMLYCAIKQMEKGPIDAITGEARYSLSEDKI	1471
	Q91823	LTERNLESNHPKILLRPESVAEKMLTNVFTFLYKFLKECAGPLFMLYCAIKQMEKGPIDAITGEARYSLSEDKI	160
110	P51805	LTERNLESNHPKILLRPESVAEKMLTNVFTFLYKFLKECAGPLFMLYCAIKQMEKGPIDAITGEARYSLSEDKI	1469

5	NOV12	IRQQIDYKTLTINCVNPEHNAPEVPVKGLICDTVTOAKEKLLDAAYKGVPSORPKAAMMLEWROGRMARITLQDELA	1597
	P70206	IRQQIDYKTLTINCVNPEHNAPEVPVKGLICDTVTOAKEKLLDAAYKGVPSORPKAAMMLEWROGRMARITLQDELA	1571
	Q9UIW2	IRQQIDYKTLTINCVNPEHNAPEVPVKGLICDTVTOAKEKLLDAAYKGVPSORPKAAMMLEWROGRMARITLQDELA	1551
	Q91823	IRQQIDYKTLTINCVNPEHNAPEVPVKGLICDTVTOAKEKLLDAAYKGVPSORPKAAMMLEWROGRMARITLQDELA	160
	P51805	IRQQIDYKTLTINCVNPEHNAPEVPVKGLICDTVTOAKEKLLDAAYKGVPSORPKAAMMLEWROGRMARITLQDELA	1548
10	NOV12	ITKILNWKRIINTLAHYQVTDGSSVALVPKQTSAYNIINNSTFTKSLRYESMLRTASSPDLRSKTPMITTDLSEGTI	1677
	P70206	ITKILNWKRIINTLAHYQVTDGSSVALVPKQTSAYNIINNSTFTKSLRYESMLRTASSPDLRSKTPMITTDLSEGTI	1651
	Q9UIW2	ITKILNWKRIINTLAHYQVTDGSSVALVPKQTSAYNIINNSTFTKSLRYESMLRTASSPDLRSKTPMITTDLSEGTI	1631
	Q91823	ITKILNWKRIINTLAHYQVTDGSSVALVPKQTSAYNIINNSTFTKSLRYESMLRTASSPDLRSKTPMITTDLSEGTI	160
	P51805	ITKILNWKRIINTLAHYQVTDGSSVALVPKQTSAYNIINNSTFTKSLRYESMLRTASSPDLRSKTPMITTDLSEGTI	1628
15	NOV12	IVHLVNHDIHLDOREGDRGSKMVSEIYITRLLATKQITLOKEVDLLETIFSTAHRGSAFLAIKYMDFLEQADKHQII	1757
	P70206	IVHLVNHDIHLDOREGDRGSKMVSEIYITRLLATKQITLOKEVDLLETIFSTAHRGSAFLAIKYMDFLEQADKHQII	1730
	Q9UIW2	IVHLVNHDIHLDOREGDRGSKMVSEIYITRLLATKQITLOKEVDLLETIFSTAHRGSAFLAIKYMDFLEQADKHQII	1710
	Q91823	IVHLVNHDIHLDOREGDRGSKMVSEIYITRLLATKQITLOKEVDLLETIFSTAHRGSAFLAIKYMDFLEQADKHQII	160
	P51805	IVHLVNHDIHLDOREGDRGSKMVSEIYITRLLATKQITLOKEVDLLETIFSTAHRGSAFLAIKYMDFLEQADKHQII	1707
20	NOV12	IVADVRHTWKSNIPLRFVWVITENPOFVFDIHNSTIDACLSVVAQTFMDSCTSEHILGKDSPEINLLYAKDITNTK	1837
	P70206	IVADVRHTWKSNIPLRFVWVITENPOFVFDIHNSTIDACLSVVAQTFMDSCTSEHILGKDSPEINLLYAKDITNTK	1809
	Q9UIW2	IVADVRHTWKSNIPLRFVWVITENPOFVFDIHNSTIDACLSVVAQTFMDSCTSEHILGKDSPEINLLYAKDITNTK	1754
	Q91823	IVADVRHTWKSNIPLRFVWVITENPOFVFDIHNSTIDACLSVVAQTFMDSCTSEHILGKDSPEINLLYAKDITNTK	160
	P51805	IVADVRHTWKSNIPLRFVWVITENPOFVFDIHNSTIDACLSVVAQTFMDSCTSEHILGKDSPEINLLYAKDITNTK	1786
25	NOV12	IVVEHRYTAKIAPISDQMSAYIAEOSRIILQNSMGATHEITSTIRYKDEVEITAAIKTEQOARORLPSKLEWV	1917
	P70206	IVVEHRYTAKIAPISDQMSAYIAEOSRIILQNSMGATHEITSTIRYKDEVEITAAIKTEQOARORLPSKLEWV	1886
	Q9UIW2	IVVEHRYTAKIAPISDQMSAYIAEOSRIILQNSMGATHEITSTIRYKDEVEITAAIKTEQOARORLPSKLEWV	1754
	Q91823	IVVEHRYTAKIAPISDQMSAYIAEOSRIILQNSMGATHEITSTIRYKDEVEITAAIKTEQOARORLPSKLEWV	160
	P51805	IVVEHRYTAKIAPISDQMSAYIAEOSRIILQNSMGATHEITSTIRYKDEVEITAAIKTEQOARORLPSKLEWV	1863
35	NOV12	IVDTMALS	1925
	P70206	IVDTMALS	1894
	Q9UIW2	IVDTMALS	1754
	Q91823	IVDTMALS	160
	P51805	IVDTMALS	1871

Tables 12E-12N list the domain descriptions from DOMAIN analysis results against NOV12. This indicates that the NOV12 sequence has properties similar to those of other proteins known to contain this domain.

Table 12E. Domain Analysis of NOV12

gnl|Smart|smart00630, Sema, semaphorin domain (SEQ ID NO:113)
CD-length = 430 residues, 100.0% aligned
Score = 242 bits (618), Expect = 1e-64

45	Query:	69	LTHLVHEQTGEVYVGAVNRIYKLSGNLTLLRAHVTGPFVEDNEKCYPPPSVQSCPHGLGS	128
	Sbjct:	1	LQNLLEDENGTLYVGARNRLVLSLNLISEAEVKTGFVLSSPDCEEC--VSKGKDPP--	56
50	Query:	129	TDNVNK-LLLLDYAANRLACGS-ASQGICQFLRLDDLFLKLGEPHHRKEHYLSSVQEAGS	186
	Sbjct:	57	TDCVNFIRLLLDYNADHLLVCGTNAFPQVCRILNLGNLDRLE-EVGRESGRGRCFPDPQHN	115
55	Query:	187	MAGVLIAGPPGQQAQKLFVGTPIID--GKSEYFPTLSSRRLMANEEDADMFGFVYQDEFVS	244
	Sbjct:	116	STAVLVLDG-----ELYVGTVAADFSGSDPAIYRSLSVRRLKGTSG-----PSLRTVL	161
60	Query:	245	SQLKIPSDTSLSKFFAFDIYVYSFRSEQFVYYLTQLQDTQLTSPDAAGEHFFTSKIVRLC	304
	Sbjct:	162	YDSRWLN-----EPNFYAFESGDFVYF---FFRETAVEDENCGKAVVSRVARVC	208
60	Query:	305	VDD-----PKFYSYVEFPIGC---EQAGVEYRLVQDAYLSRPGRAIAHQLGLAEDD	353
	Sbjct:	209	KNDVGGPRSLSKKWTSLFKARLECSVPGEFFPYFNLQAAFLLPAG-----SESDD	259
60	Query:	354	VLFTVFAQQGQKNRVKPPKESALCFLTLRAIKIKERIQSCYRGEGLSL----PWLNLK	409
	Sbjct:	260	VLYGVFSTS----SNPIPGSAVCAFSLSDINAVFNEPFKECETGNSQWLPYPRGLVVFPR	315

Query:	410	ELGCINSPLQI---DDDFC-GQDFNQPLGGTVTIEGTPLFV--DKDDGLTAVAA---Y	458
		+ + +	
Sbjct:	316	PGTCPNTPLSSKDLPPDVLNFIKTHPLMDEVVQPLTGRPLFKVKTDSNYLLTSAIVDRVRT	375
Query:	459	DYRGRTVVFAGTRSGRIRKILVDLSNPGGRPALAYESVVAQEGSPILRDVLVSPNH	514
		+ +	
Sbjct:	376	DGGNYTVLFGTSDGRILKVVLSSSSSESVLEEISVFDGPGSPV-SDLVLSPKK	430

gnl|Pfam|pfam01403, Sema, Sema domain. The Sema domain occurs in semaphorins, which are a large family of secreted and transmembrane proteins, some of which function as repellent signals during axon guidance. Sema domains also occur in the hepatocyte growth factor receptor. (SEQ ID NO:114)
CD-Length = 433 residues, 99.5% aligned
Score = 171 bits (432), Expect = 5e-43

Query:	69	LTHLVVHEQTGEVYVGAVNRIYKLSGN----LTLRAHVTGPVEDNEKCYPPPSVQSCPH	124
		++ + + + + + +	
Sbjct:	1	FVTLLDDEDRGRLYVGARNRVYVLNLEDLSEVLNLKTGWPGSCETCEECNMKGKSP----	56
Query:	125	GLGSTDNVN-KLLLLDYAANRLLAGCS-ASQGCQFLRLDDLEKLGEPHHRKEHYLSSVQ	182
		+ + + + + +	
Sbjct:	57	---LTECTNFIRVLQAYNDTHLYVCGTNAFQPVCTLINLGLDLSLDVDNDEEDGCGDCPYD	113
Query:	183	EAGSMAGVLIAGPPGQGAKLFVGTPIDGKSEYFPTLSSRRIMANEEDADMFGFVYQDEF	242
		+ + + + + + + +	
Sbjct:	114	PLGNTTSLVLVQG-----GELYSGTVID-----FSGRDPISIRRLGSHDGLRTEFHD--	159
Query:	243	VSSQLKIPSDTSLSKFFAPFDIYVYSFRSEQFVYYLTLQLDTQLTSPDAAGEHFFTSKIVR	302
		+ + ++ + + + + + + + ++	
Sbjct:	160	-SKWLNLPNFVD----SYPIHYVHSF-SDDKVYF----FFRETAVEDSNCKTIH-SRVAR	208
Query:	303	LCVDDPKFYSYVEFPIGC-----EQAGVEYRLVQDAYLSRPGRALAHQLGLA	349
		+ + + + + +	
Sbjct:	209	VCKNDPGGRSYLELNKWTTFILKARLNCSIPGEGTPFYFNELQAFAVLPTG-----A	259
Query:	350	EDEDVLFVTFQAQGQKNRVKPKESALCLFTLRAIKE--KIKERIQSCYRGEGLSLPWLL	407
		+ + + + + + + +	
Sbjct:	260	DTDPVLVYGVFTTS----SNSSAGSAVCAFSMSDINQVFEGEPFKHQSPNSKWLPYRGKVPQ	315
Query:	408	NKELGCINSP-LQIDDDFCGQDFNQPLGGTGT--IEGTPLFVDKDDG--LTAVA-----A	457
		+ + + + + +++	
Sbjct:	316	PRPGQCPNASGLNLPDDTLNFIRCHPLMDEVVPPLHNVPLEFVGQSGNYRLTSAIVDRVRA	375
Query:	458	YDYRGRTVVFAGTRSGRIRKILVDLSNPGGR---PALAYESVVAQEGSPILRDLVLS	511
		+ + + + + + + ++	
Sbjct:	376	GDGQIYTVLFLGTDDGRV-LKQVVLRSRSSASYLVVVLESLSVFPDGEVPQRMVSS	431

gnl|Pfam|pfam01833, TIG, IPT/TIG domain. This family consists of a domain that has an immunoglobulin like fold. These domains are found in cell surface receptors such as Met and Ron as well as in intracellular transcription factors where it is involved in DNA binding. (SEQ ID NO:115)

CD-Length = 85 residues, 100.0% aligned
Score = 78.2 bits (191), Expect = 4e-15

Query: 983 PTFYRVSPSRGPLSGGTWIGIEGSHLNAGSDVAVSVGGRPCSFWSRRNSREIRCLTPPG 1042
 | +||| ||||| | | ||+| +| |+| +| | | + + +| | |||
 Sbjct: 1 PVITSISPSSGPLSGGTEITITGSNLGSGEDIKVTFGGTECDV--VSQEASQIVCKTPPY 58

5 Query: 1043 QSPGSAPIIININRAQLTNPEVKYNYT 1069
 + | |+ ++++ |++ | + |
 Sbjct: 59 ANGGPQPVTVSLDGGGLSSSPVTFTYV 85

Table 12H. Domain Analysis of NOV12

gnl|Pfam|pfam01833, TIG, IPT/TIG domain. This family consists of a domain that has an immunoglobulin like fold. These domains are found in cell surface receptors such as Met and Ron as well as in intracellular transcription factors where it is involved in DNA binding. (SEQ ID NO:115)
 CD-Length = 85 residues, 100.0% aligned
 Score = 60.1 bits (144), Expect = 1e-09

10

Query: 886 PKILKLSPETGPRQGGTRLTITGENLGLRFEDVRLGVRVGKVLCSPESEYISAEQIVCE 945
 | | +|| +|| ||| +||| ||| + | | | | | |||+
 Sbjct: 1 PVITSISPSSGPLSGGTEITITGSNLGS---GEDIKVTFGGTECDVVSQEA---SQIVCK 54

15 Query: 946 IGDASSVRAHDALVEVCVRDCSPHYRALSPKRFTEV 981
 ++ | + | || ||+|
 Sbjct: 55 TPPYANGGPQPVTVSLDGGGLSS-----SPVTFTYV 85

Table 12I. Domain Analysis of NOV12

gnl|Pfam|pfam01833, TIG, IPT/TIG domain. This family consists of a domain that has an immunoglobulin like fold. These domains are found in cell surface receptors such as Met and Ron as well as in intracellular transcription factors where it is involved in DNA binding. (SEQ ID NO:115)
 CD-Length = 85 residues, 100.0% aligned
 Score = 46.6 bits (109), Expect = 1e-05

20 Query: 1173 PVLEPLSPTGLLELKPSSPLILKGRNLLPPAPGNSRLNYTVLIGSTPCTLT-VSETQLLC 1231
 ||+ +||+ | + + + | || | + | | | | + +|++|
 Sbjct: 1 PVITSISPSSG-PLSGGTEITITGSNL-----GSGEDIKVTFGGTECDVVSQEASQIVC 53

25 Query: 1232 EAPNLTGQH----KVTVRAGGFESPGTLQVY 1259
 + | |++ || || |
 Sbjct: 54 KTPPYANGGPQPVTVSLDGGGLSSSPVTFTYV 85

Table 12J. Domain Analysis of NOV12

gnl|Smart|smart00429, IPT, ig-like, plexins, transcription factors (SEQ ID NO:116)
 CD-Length = 93 residues, 100.0% aligned
 Score = 70.9 bits (172), Expect = 6e-13

30 Query: 885 DPKILKLSPETGPRQGGTRLTITGENLGLRFEDVRLGVRVGKVLCSPESEYISAEQIVC 944
 || | ++|| +|| |||+|+ |+|| | + | ||+| |+ + |+ | |||
 Sbjct: 1 DPVITRISPNSGPLSGGTRITLCGKNLDS-ISVVFVEVGVEVPCTFLPSDV-SQTAIVC 58

Query: 945 EIGDASSVRAHDALVEVCVRDCSPHYRALSPKRFTFV 981
 + | | | + | |||
 Sbjct: 59 KTP-PYHNIPGSVPVRVEVGLRNGGVPG-EPSPFTYV 93

5

Table 12K. Domain Analysis of NOV12

gnl|Pfam|pfam01437, Plexin repeat, Plexin repeat. A cysteine rich repeat found in several different extracellular receptors. The function of the repeat is unknown. Three copies of the repeat are found Plexin. Two copies of the repeat are found in mahogany protein. A related *C. elegans* protein contains four copies of the repeat. The Met receptor contains a single copy of the repeat. The Pfam alignment shows 6 conserved cysteine residues that may form three conserved disulphide bridges. (SEQ ID NO:117)
 CD-Length = 48 residues, 100.0% aligned
 Score = 59.3 bits (142), Expect = 2e-09

10 Query: 532 SCVQYTSCELCLGSRDPHCGWCVLHSMCSRRDACERADEPQRFADLLQCV 582
 +| | ||| || + || ||| | + | | + ++ |
 Sbjct: 1 NCSQHTSCGSCLSAPDPGCGWCPSPKRKCTRLEECR---GEGWSQSQETCP 48

Table 12L. Domain Analysis of NOV12

gnl|Pfam|pfam01437, Plexin repeat, Plexin repeat. A cysteine rich repeat found in several different extracellular receptors. The function of the repeat is unknown. Three copies of the repeat are found Plexin. Two copies of the repeat are found in mahogany protein. A related *C. elegans* protein contains four copies of the repeat. The Met receptor contains a single copy of the repeat. The Pfam alignment shows 6 conserved cysteine residues that may form three conserved disulphide bridges. (SEQ ID NO:117)
 CD-Length = 48 residues, 100.0% aligned
 Score = 53.5 bits (127), Expect = 1e-07

15 Query: 681 NCSVHQSSCLSCVNGSFP-CHWCKYRHVCTHNVADCAFLEGRVNVSEDCP 729
 ||| | | ||++ | | | | | +|+ || | ||
 Sbjct: 1 NCSQHTS-CGSCLSAPDPGCGWCPSPKRKCTRL-EECSRGEWSQSQETCP 48

Table 12M. Domain Analysis of NOV12

gnl|Pfam|pfam01437, Plexin repeat, Plexin repeat. A cysteine rich repeat found in several different extracellular receptors. The function of the repeat is unknown. Three copies of the repeat are found Plexin. Two copies of the repeat are found in mahogany protein. A related *C. elegans* protein contains four copies of the repeat. The Met receptor contains a single copy of the repeat. The Pfam alignment shows 6 conserved cysteine residues that may form three conserved disulphide bridges. (SEQ ID NO:117)
 CD-Length = 48 residues, 89.6% aligned
 Score = 46.2 bits (108), Expect = 2e-05

20

Query: 835 RESCGLCLKADPRFECGWCVAERRCSLRHHCAADTPASWMHARHGSSRC 883
 ||| || | | ||| + +||+ | +
 Sbjct: 5 HTSCGSCLSA-PDPGCGWCPSRKRCRTRLEEC-----SRGEGWSQSQETC 47

Table 12N. Domain Analysis of NOV12

gnl|Smart|smart00423, PSI, domain found in Plexins, Semaphorins and Integrins (SEQ ID NO:118)
 CD-Length = 47 residues, 89.4% aligned
 Score = 44.3 bits (103), Expect = 6e-05

Query: 833 ALRESCGLCLKADPRFECGWCVAERRCSLRHHCAADTPASWMHA 876
 + || || | + | || ++ ||+ | + +|
 Sbjct: 3 SAYTSCSECLLARDPY-CAWCSSQGRCTSGERCDS-LRQNWSSG 44

5
 10 Plexin is a type I membrane protein which was identified in *Xenopus* nervous system by hybridoma technique. Molecular cloning studies demonstrated that the extracellular segment of the plexin protein possesses three internal repeats of cysteine cluster which are homologous to the cysteine-rich domain of the c-met proto-oncogene protein product. A cell aggregation test revealed that the plexin protein mediated cell adhesion via a homophilic
 15 binding mechanism, in the presence of calcium ions. Plexin was expressed in the neuronal elements composing particular neuron circuits in *Xenopus* CNS and PNS. These findings indicate that plexin is a new member of the Ca(2+)-dependent cell adhesion molecules, and suggest that the molecule plays an important role in neuronal cell contact and neuron network formation.

20 In the developing nervous system axons navigate with great precision over large distances to reach their target areas. Chemorepulsive signals such as the semaphorins play an essential role in this process. The effects of one of these repulsive cues, semaphorin 3A (Sema3A), are mediated by the membrane protein neuropilin-1 (Npn-1). Recent work has shown that neuropilin-1 is essential but not sufficient to form functional Sema3A receptors
 25 and indicates that additional components are required to transduce signals from the cell surface to the cytoskeleton. Members of the plexin family interact with the neuropilins and act as co-receptors for Sema3A. Neuropilin/plexin interaction restricts the binding specificity of neuropilin-1 and allows the receptor complex to discriminate between two different semaphorins. Deletion of the highly conserved cytoplasmic domain of Plexin-A1 or -A2
 30 creates a dominant negative Sema3A receptor that renders sensory axons resistant to the repulsive effects of Sema3A when expressed in sensory ganglia. These data suggest that functional semaphorin receptors contain plexins as signal-transducing and neuropilins as ligand-binding subunits.

Physiologic SEMA3A receptors consist of NRP1/PLXN1 complexes. Two semaphorin-binding proteins, plexin-1 (PLXN1) and neuropilin-1 (NRP1; 602069), form a stable complex. While SEMA3A binding to NRP1 does not alter nonneuronal cell morphology, SEMA3A interaction with NRP1/PLXN1 complexes induces adherent cells to round up. Expression of a dominant-negative PLXN1 in sensory neurons blocked SEMA3A-induced growth cone collapse. SEMA3A treatment led to the redistribution of growth cone NRP1 and PLXN1 into clusters.

The semaphorin family of proteins constitute one of the major cues for axonal guidance. The prototypic member of this family is Sema3A, previously designated semD/III or collapsin-1. Sema3A acts as a diffusible, repulsive guidance cue in vivo for the peripheral projections of embryonic dorsal root ganglion neurons. Sema3A binds with high affinity to neuropilin-1 on growth cone filopodial tips. Although neuropilin-1 is required for Sema3A action, it is incapable of transmitting a Sema3A signal to the growth cone interior. Instead, the Sema3A/neuropilin-1 complex interacts with another transmembrane protein, plexin, on the surface of growth cones. Certain semaphorins, other than Sema3A, can bind directly to plexins. The intracellular domain of plexin is responsible for initiating the signal transduction cascade leading to growth cone collapse, axon repulsion, or growth cone turning. This intracellular cascade involves the monomeric G-protein, Rac1, and a family of neuronal proteins, the CRMPs. Rac1 is likely to be involved in semaphorin-induced rearrangements of the actin cytoskeleton, but how plexin controls Rac1 activity is not known. Vertebrate CRMPs are homologous to the *Caenorhabditis elegans* unc-33 protein, which is required for proper axon morphology in worms. CRMPs are essential for Sema3A-induced, neuropilin-plexin-mediated growth cone collapse, but the molecular interactions of growth cone CRMPs are not well defined. Mechanistic aspects of plexin-based signaling for semaphorin guidance cues may have implications for other axon guidance events and for the basis of growth cone motility.

In *Drosophila*, plexin A is a functional receptor for semaphorin-1a. The human plexin gene family comprises at least nine members in four subfamilies. Plexin-B1 is a receptor for the transmembrane semaphorin Sema4D (CD100), and plexin-C1 is a receptor for the GPI-anchored semaphorin Sema7A (Sema-K1). Secreted (class 3) semaphorins do not bind directly to plexins, but rather plexins associate with neuropilins, coreceptors for these semaphorins. Plexins are widely expressed: in neurons, the expression of a truncated plexin-A1 protein blocks axon repulsion by Sema3A. The cytoplasmic domain of plexins associates with a tyrosine kinase activity. Plexins may also act as ligands mediating repulsion in epithelial cells

in vitro. Thus, plexins are receptors for multiple (and perhaps all) classes of semaphorins, either alone or in combination with neuropilins, and trigger a novel signal transduction pathway controlling cell repulsion.

In addition, recent studies have identified semaphorins and their receptors as putative molecular cues involved in olfactory pathfinding, plasticity and regeneration. The semaphorins comprise a large family of secreted and transmembrane axon guidance proteins, being either repulsive or attractive in nature. Neuropilins were shown to serve as receptors for secreted class 3 semaphorins, whereas members of the plexin family are receptors for class 1 and V (viral) semaphorins.

The disclosed NOV12 nucleic acid of the invention encoding a Plexin-1-like protein includes the nucleic acid whose sequence is provided in Table 12A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 12A while still encoding a protein that maintains its Plexin-1-like activities and physiological functions, or a fragment of such a nucleic acid.

The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 29 percent of the bases may be so changed.

The disclosed NOV12 protein of the invention includes the Plexin-1-like protein whose sequence is provided in Table 12B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 12B, while still encoding a protein that maintains its Plexin-1-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 29 percent of the residues may be so changed.

The protein similarity information, expression pattern, and map location for the plexin-1-like protein and the NOV12 protein disclosed herein suggest that this plexin-1-like protein may have important structural and/or physiological functions characteristic of the mannosidase

protein family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These applications include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon.

The NOV12 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from AIDS, cancer therapy, treatment of Neurologic diseases, Brain and/or autoimmune disorders like encephalomyelitis, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, endocrine diseases, muscle disorders, inflammation and wound repair, bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome, and/or other pathologies/disorders. The NOV12 nucleic acid, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV12 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV12 protein have multiple hydrophilic regions, each of which can be used as an immunogen. This novel protein also has value in development of powerful assay system for functional analysis of various human

disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOVX Nucleic Acids and Polypeptides

5 One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term
10 “nucleic acid molecule” is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

15 An NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a “mature” form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the
20 polypeptide, precursor or proprotein encoded by an ORF described herein. The product “mature” form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps as they may take place within the cell, or host cell, in which the gene product arises. Examples of such processing steps leading to a “mature” form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded
25 by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal
30 sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a “mature” form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation,

myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probes", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as utilized herein, is one, which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, or a complement of this aforementioned nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31 as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard

PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term “oligonucleotide” refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue.

Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of an NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, or 31 is one that is sufficiently complementary to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, or 31 that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, thereby forming a stable duplex.

As used herein, the term “complementary” refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term “binding” means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the

effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. *See e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for an NOVX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide

sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human NOVX protein.

Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

An NOVX polypeptide is encoded by the open reading frame ("ORF") of an NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, e.g., a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, e.g. from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, or 31; or an anti-sense strand nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, or 31; or of a naturally occurring mutant of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express an NOVX protein, such as by measuring a level of an NOVX-encoding nucleic acid in a sample of cells from a subject e.g., detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of an NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, or 31, that encodes a polypeptide having an NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

NOVX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31 due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32.

In addition to the human NOVX nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding an NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from the human SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of

the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

5 Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet 10 another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other 15 than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no 20 other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at 25 which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at T_m , 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at 30 pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequences SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well-known within the art. *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY

MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* 78: 6789-6792.

Conservative Mutations

5 In addition to naturally-occurring allelic variants of NOVX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, thereby leading to changes in the amino acid sequences of the encoded NOVX proteins, without altering the functional ability of said NOVX proteins. For example, 10 nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the NOVX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For 15 example, amino acid residues that are conserved among the NOVX proteins of the invention are predicted to be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such 20 NOVX proteins differ in amino acid sequence from SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 45% homologous to the amino acid sequences SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, and 32. 25 Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, and 32; more preferably at least about 70% homologous SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32; still more preferably at least about 80% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32; even more preferably at least about 90% homologous 30 to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32; and most preferably at least about 95% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32.

An isolated nucleic acid molecule encoding an NOVX protein homologous to the protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32 can be

created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

5 Mutations can be introduced into SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue
10 having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine,
15 methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an NOVX coding sequence, such as by saturation
20 mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain
25 interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one
30 of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, VLIM, HFY, wherein the letters within each group represent the single letter amino acid code.

In one embodiment, a mutant NOVX protein can be assayed for (i) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically-active portions thereof, (ii) complex formation between a mutant NOVX protein

and an NOVX ligand; or (iii) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically-active portion thereof; (e.g. avidin proteins).

In yet another embodiment, a mutant NOVX protein can be assayed for the ability to regulate a specific biological function (e.g., regulation of insulin release).

5 **Antisense Nucleic Acids**

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a
10 nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules
15 encoding fragments, homologs, derivatives and analogs of an NOVX protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32, or antisense nucleic acids complementary to an NOVX nucleic acid sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding
20 region" of the coding strand of a nucleotide sequence encoding an NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences which
25 flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary
30 to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15,

20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified
5 nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine,
10 xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil,
15 beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the
20 antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a
25 subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in
30 the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface

(e.g., by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other. See, e.g., Gaultier, *et al.*, 1987. *Nucl. Acids Res.* 15: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (See, e.g., Inoue, *et al.* 1987. *Nucl. Acids Res.* 15: 6131-6148) or a chimeric RNA-DNA analogue (See, e.g., Inoue, *et al.*, 1987. *FEBS Lett.* 215: 327-330).

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for an NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of an NOVX cDNA disclosed herein (*i.e.*, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an NOVX-encoding mRNA. See, e.g., U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. *See, e.g.*, Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g.*, Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S₁ nucleases (*See*, Hyrup, *et al.*, 1996. *supra*); or as probes or primers for DNA sequence and hybridization (*See*, Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.*, 1996. *supra*).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (*see*, Hyrup, *et al.*, 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.*, 1996.

supra and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.*, Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.*, Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32 while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, an NOVX variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed

by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, an NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX proteins having less than about 30% (by dry weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX proteins, still more preferably less than about 10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the NOVX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of an NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically-active portion of an NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence shown SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32, and retains the functional activity of the protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32, and retains the functional activity of the NOVX proteins of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known

in the art, such as GAP software provided in the GCG program package. *See*, Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, an NOVX "chimeric protein" or "fusion protein" comprises an NOVX polypeptide operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an NOVX protein SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within an NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of an NOVX protein. In one embodiment, an NOVX fusion protein comprises at least one biologically-active portion of an NOVX protein. In another embodiment, an NOVX fusion protein comprises at least two biologically-active portions of an NOVX protein. In yet another embodiment, an NOVX

fusion protein comprises at least three biologically-active portions of an NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX polypeptides.

In another embodiment, the fusion protein is an NOVX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of NOVX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is an NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between an NOVX ligand and an NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of an NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with an NOVX ligand.

An NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be

carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g., Ausubel, et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992*). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g., a GST polypeptide*). An NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

NOVX Agonists and Antagonists

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e., mimetics*) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g., discrete point mutation or truncation of the NOVX protein*). An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

Variants of the NOVX proteins that function as either NOVX agonists (*i.e., mimetics*) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g., truncation mutants*) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g., for phage display*) containing the set of NOVX sequences therein. There are a variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate

set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.*; Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; 5 Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent 10 selection of variants of an NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single 15 stranded portions from reformed duplexes by treatment with S₁ nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of combinatorial 20 libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, 25 transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants. 30 *See, e.g.*, Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

Anti-NOVX Antibodies

Also included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that
5 contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} , F_{ab}' and $F_{(ab)2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes
10 have subclasses as well, such as IgG₁, IgG₂, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated NOVX-related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to
15 generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein and encompasses an epitope thereof such that an
20 antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these
25 are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX-related protein that is located on the surface of the protein, e.g., a hydrophilic region. A hydrophobicity analysis of the human NOVX-related protein sequence will indicate which regions of a NOVX-related protein are particularly
30 hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, e.g., Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78:

3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

5 A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives,
10 fragments, analogs homologs or orthologs thereof (see, for example, *Antibodies: A Laboratory Manual*, Harlow and Lane, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

Polyclonal Antibodies

15 For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a
20 recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include,
25 but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and *Corynebacterium parvum*, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant
30 (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific

antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, *Nature*, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, *J. Immunol.*, 133:3001 (1984); Brodeur et al., MONOCLONAL ANTIBODY PRODUCTION TECHNIQUES AND APPLICATIONS, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, *Anal. Biochem.*, 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to

obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, *Nature* 368, 812-13 (1994)) or by covalently joining to the

immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')₂ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., *Nature*, 321:522-525 (1986); Riechmann et al., *Nature*, 332:323-327 (1988); Verhoeven et al., *Science*, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, *Curr. Op. Struct. Biol.*, 2:593-596 (1992)).

Human Antibodies

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 *Immunol Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. *Proc Natl Acad Sci USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, *J. Mol. Biol.*, 227:381 (1991); Marks et al., *J. Mol. Biol.*, 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (*Bio/Technology* 10, 779-783 (1992)); Lonberg et al. (*Nature* 368 856-859 (1994)); Morrison (*Nature* 368, 812-13 (1994)); Fishwild et al. (*Nature Biotechnology* 14, 845-51 (1996)); Neuberger (*Nature Biotechnology* 14, 826 (1996)); and Lonberg and Huszar (*Intern. Rev. Immunol.* 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full

complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins.

The antibodies can be obtained directly from the animal after immunization with an

immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective

identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an F_{(ab)₂} fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an F_{(ab)₂} fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, *Nature*, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh *et al.*, *Methods in Enzymology*, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part

of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. F(ab')₂ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., *Science* 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate F(ab')₂ fragments. These fragments are reduced in the presence of the dithiol com26S protease regulatory subunit 4g agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., *J. Exp. Med.* 175:217-225 (1992) describe the production of a fully humanized bispecific antibody F(ab')₂ molecule. Each Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., *J. Immunol.* 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., *Proc. Natl. Acad. Sci. USA* 90:6444-6448 (1993) has

provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain.

Accordingly, the V_H and V_L domains of one fragment are forced to pair with the

5 complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., *J. Immunol.* 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., *J. Immunol.* 147:60 (1991).

10 Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (Fc γ R), such as Fc γ RI (CD64), Fc γ RII (CD32) and Fc γ RIII (CD16) so as
15 to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds
20 tissue factor (TF).

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such
25 antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by
30 forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., *J. Exp Med.*, 176: 1191-1195 (1992) and Shopes, *J. Immunol.*, 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. *Cancer Research*, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., *Anti-Cancer Drug Design*, 3: 219-230 (1989).

Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi , ^{131}I , ^{131}In , ^{90}Y , and ^{186}Re .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates

(such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such as streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of an NOVX protein is facilitated by generation of hybridomas that bind to the fragment of an NOVX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an NOVX protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Anti-NOVX antibodies may be used in methods known within the art relating to the localization and/or quantitation of an NOVX protein (e.g., for use in measuring levels of the NOVX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for NOVX proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds (hereinafter "Therapeutics").

An anti-NOVX antibody (e.g., monoclonal antibody) can be used to isolate an NOVX polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-NOVX antibody can facilitate the purification of natural NOVX polypeptide from cells and of recombinantly-produced NOVX polypeptide expressed in host cells. Moreover, an anti-NOVX antibody can be used to detect NOVX protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the NOVX protein. Anti-NOVX antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the

antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid

sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition